

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:36:12 ; Search time 84 Seconds  
(without alignments)  
1219.112 Million cell updates/sec

Title: CAC80065  
Perfect score: 2576  
Sequence: 1 MSALVLRCAVAVSIRGSSCR.....HELLHTHYVAGVEKDEKK 497

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2576	100.0	497	10 Q9FYU1	Q9FYU1 chlamydomon
2	1702.5	66.1	505	10 Q8VZ20	Q8VZ20 chlamydomon
3	1409	54.7	436	10 Q8VX03	Q8VX03 chlamydomon
4	1384.5	53.7	449	10 Q9AR66	Q9AR66 chlamydomon
5	1196	46.4	403	10 Q9AU60	Q9AU60 chlamydomon
6	1010.5	39.2	608	16 Q9WY44	Q9WY44 thermotoga
7	935.5	36.3	581	16 Q8B8C8	Q8B8C8 thermotoga
8	903.5	35.1	579	2 Q9XC55	Q9XC55 clostridium
9	893	34.7	555	3 Q8TG63	Q8TG63 piromyces s
10	880.5	34.2	585	3 Q8TFP2	Q8TFP2 neocallimastix
11	877	34.0	585	2 Q46508	Q46508 desulfovibrio
12	874	32.0	1206	5 Q96948	Q96948 neocallimastix
13	816.5	31.7	578	2 Q93SF7	Q93SF7 eubacterium
14	805.5	31.3	589	5 Q9GTP2	Q9GTP2 trichomonas
15	794.5	30.8	606	2 Q46606	Q46606 desulfovibrio
16	778.5	30.2	484	2 Q9RCN3	Q9RCN3 megasphaera

17	741	28.8	582	16 Q59262	Q59262 clostridium
18	738.5	28.7	468	5 Q27094	Q27094 trichomonas
19	737.5	28.6	574	2 Q59261	Q59261 clostridium
20	732.5	28.4	572	2 Q92N84	Q92N84 clostridium
21	730.5	28.4	572	16 Q9RH08	Q9RH08 clostridium
22	659	25.6	421	2 Q08311	Q08311 desulfovibrio
23	652	25.3	449	5 Q27096	Q27096 trichomonas
24	590.5	22.9	421	2 Q9AM36	Q9AM36 desulfovibrio
25	555	21.5	467	5 Q9GTP1	Q9GTP1 spirochaeta
26	512.5	19.9	645	16 Q52683	Q52683 thermotoga
27	504.5	19.6	464	5 Q9GTP3	Q9GTP3 entamoeba h
28	504.5	19.6	468	5 Q9GTX0	Q9GTX0 entamoeba h
29	492.5	19.1	476	11 Q9D320	Q9D320 mus musculus
30	489.5	19.0	476	11 Q9CX86	Q9CX86 mus musculus
31	478	18.6	474	10 Q9AC16	Q9AC16 arabidopsis
32	465.5	18.1	369	5 Q8T5G1	Q8T5G1 trichomonas
33	461.5	17.9	476	10 Q8W303	Q8W303 oryza sativa
34	459	17.8	479	10 Q93YF9	Q93YF9 medicago tr
35	457	17.7	456	5 Q9TY15	Q9TY15 neocallimastix
36	456.5	17.7	476	4 Q9H6Q4	Q9H6Q4 homo sapiens
37	456.5	17.7	525	4 Q96S10	Q96S10 homo sapiens
38	451	17.5	369	5 Q8T6P9	Q8T6P9 trichomonas
39	446.5	17.3	369	5 Q8T5G2	Q8T5G2 trichomonas
40	445	17.3	474	5 Q9BKJ3	Q9BKJ3 glardia lam
41	437.5	17.0	374	4 Q9H6J8	Q9H6J8 homo sapiens
42	390.5	15.2	462	11 Q9CY07	Q9CY07 mus musculus
43	388.5	15.1	477	5 Q8SY87	Q8SY87 drosophila
44	367	14.2	457	5 Q9N392	Q9N392 caenorhabditis
45	361.5	14.0	456	4 Q9UHQ1	Q9UHQ1 homo sapiens

## ALIGNMENTS

RESULT 1  
ID Q9FYU1 PRELIMINARY; PRT; 497 AA.  
Q9FYU1:  
AC 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Fe-hydrogenase precursor (EC 1.18.99.1) (Iron-hydrogenase HydA).  
GN HYD1 OR HYDA.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kaminski A.U., Happe T.;  
RC "Isolation and characterization of the hydA gene encoding the Fe-hydrogenase of Chlamydomonas reinhardtii.";  
RT Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Forestier M., Zhang L., Plummer S., Ahmann D., Selbert M., Ghirelli M.L.;  
RC "Two putative Fe-only hydrogenases cloned from Chlamydomonas reinhardtii are coexpressed in cells undergoing anaerobiosis.";  
RT Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Forestier M., Zhang L., Plummer S., Ahmann D., Selbert M., Ghirelli M.L.;  
RC "Two putative Fe-only hydrogenases cloned from Chlamydomonas reinhardtii are coexpressed in cells undergoing anaerobiosis.";  
RT Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF288201; AAG00591.1; -;  
DR EMBL; AF288201; AAG00591.1; -;  
DR EMBL; AF288201; AAG00591.1; -;  
DR HSP; P29166; 1FEH.

DR InterPro: IPR004108; Fe\_hyd\_19\_C.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_19\_C.1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU.1.  
 DR Oxidoreductase; Transl. peptidase.  
 FT TRANSIT 1  
 FT CHAIN 57  
 SQ SEQUENCE 497 AA; 53112 MW; 2E618A259E6572F4 CRC64;

Query Match 100.0%; Score 2576; DB 10; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 2e-175;  
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSALVLRKCAVSVIRGSSCARQVAPRAPLAATVRAVLAATLEAPARRLGNAVCAAAAPA 60  
 DB 1 MSALVLRKCAVSVIRGSSCARQVAPRAPLAATVRAVLAATLEAPARRLGNAVCAAAAPA 60  
 DB 61 AEAFLSHVQOALAEALAPKDDPTKHKVCVAPAVVAIAETLGLAAGATTPKQLAEGLR 120  
 QY 121 RLGEDEVFTLFGADLTMEGSELRLHLEHAPHSDPLPMFTSCCPGWTAMLEKS 180  
 DB 121 RLGEDEVFTLFGADLTMEGSELRLHLEHAPHSDPLPMFTSCCPGWTAMLEKS 180  
 QY 181 YPDLIPIYVSSCKSPQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVD 240  
 DB 181 YPDLIPIYVSSCKSPQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVD 240  
 QY 241 ADPLRLDHLVITVELGNIFKEKGINLAELPECEMDNPMGVSGAGVLTGGVGEAA 300  
 DB 241 ADPLRLDHLVITVELGNIFKEKGINLAELPECEMDNPMGVSGAGVLTGGVGEAA 300  
 QY 301 LRTAYELFTGTPLRLSLSEVRGMDGKETNTMVPAPGSKFEBELKRAAAAEAAAHG 360  
 DB 301 LRTAYELFTGTPLRLSLSEVRGMDGKETNTMVPAPGSKFEBELKRAAAAEAAAHG 360  
 QY 361 TPGPLAMDGAGFTSEDRGCGITLRAVANGLGNAKKLITKMOAGEAKYDFEIMACPAQ 420  
 DB 361 TPGPLAMDGAGFTSEDRGCGITLRAVANGLGNAKKLITKMOAGEAKYDFEIMACPAQ 420  
 QY 421 CVGGGGOPRSTDKATQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPLGKHAHEL 480  
 DB 421 CVGGGGOPRSTDKATQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPLGKHAHEL 480  
 QY 481 LHTHYVAGVGEKDEKK 497  
 DB 481 LHTHYVAGVGEKDEKK 497

RESULT 2  
 Q8VZ20 PRELIMINARY; PRT; 505 AA.  
 AC Q8VZ20;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Iron-hydrogenase HydB.  
 GN HYDB.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=21GR;  
 RA Forestier M., Zhang L., Plummer S., Ahmann D., Seibert M.,  
 RA Chirardi M.L.;  
 RT "Two putative Fe-only hydrogenases cloned from Chlamydomonas  
 reinhardtii are coexpressed in cells undergoing anaerobiosis";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY055756; MAL23573.1;  
 DR InterPro: IPR004108; Fe\_hyd\_19\_C.

DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_19\_C.1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU.1.  
 DR Oxidoreductase; Transl. peptidase.  
 FT TRANSIT 1  
 FT CHAIN 57  
 SQ SEQUENCE 505 AA; 53121 MW; 3A1948749B033EA9 CRC64;

Query Match 66.1%; Score 1702.5; DB 10; Length 505;  
 Best Local Similarity 67.7%; Pred. No. 3.7e-113;  
 Matches 331; Conservative 57; Mismatches 84; Indels 17; Gaps 5;

QY 18 SCARQVAPRAPLAATVRAVLA-----TLEAPARRLGNAVCAAAAEAPLSHYQA 71  
 DB 16 ACARRTNAPAHPAVAPVCLPSRAGKFFNLQKVPSSQASGSTRIVATATDAVPHKLA 75  
 QY 72 LAELAPKDDPTKHKVCVAPAVVAIAETLGLAAGATTPKQLAEGLR 131  
 DB 72 LAELAPKDDPTKHKVCVAPAVVAIAETLGLAAGATTPKQLAEGLR 131  
 QY 132 FGADLTMEGSELRLHLEHAPHSDPLPMFTSCCPGWTAMLEKSTPDLIPYVSSC 191  
 DB 135 FAADLTMEGSELRLHLEHAPHSDPLPMFTSCCPGWTAMLEKSTPDLIPYVSSC 194  
 QY 192 KSPQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVDADPTLRDHLV 251  
 DB 195 KSPQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVDADPTLRDHLV 253  
 QY 252 ITTVELGNIFKEKGINLAELPECEMDNPMGVSGAGVLTGGVGEAA 311  
 DB 254 ITTVELGNIFKEKGINLAELPECEMDNPMGVSGAGVLTGGVGEAA 313  
 QY 312 PLRLSLSEVRGMDGKETNTMVPAPGSKFEBELKRAAAAEAAAHG-TP 362  
 DB 314 PLRLSLSEVRGMDGKETNTMVPAPGSKFEBELKRAAAAEAAAHG-TP 373  
 QY 363 GPLAMGAGFTSEDRGCGITLRAVANGLGNAKKLITKMOAGEAKYDFEIMACPAQ 422  
 DB 374 GPLAMGAGFTSEDRGCGITLRAVANGLGNAKKLITKMOAGEAKYDFEIMACPAQ 433  
 QY 423 GGGGGOPRSTDKATQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPLGKHAHEL 482  
 DB 434 GGGGGOPRSTDKATQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPLGKHAHEL 493  
 QY 483 THYVAGVE 491  
 DB 494 THYVAGVE 502

RESULT 3  
 Q8VX03 PRELIMINARY; PRT; 436 AA.  
 AC Q8VX03;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Fe-hydrogenase precursor (EC 1.18.99.1).  
 GN HYDB.  
 OS Chlorella fusca.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales;  
 OC Scenedesmeaceae; Scenedesmus.  
 OX NCBI\_TaxID=3073;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Happe T.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Winkler M.;  
 RA "The Fe-hydrogenase of Chlorella fusca";  
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ298228; CAC83291.1;  
 DR InterPro: IPR004108; Fe\_hyd\_19\_C.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_19\_C.1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU.1.

KW Transit peptide: Oxidoreductase.  
FT TRANSIT 1 21 POTENTIAL.  
FT CHAIN 22 436 FE-HYDROGENASE.  
SQ SEQUENCE 436 AA; 47318 MW; BF7299732FBI984 CRC64;

Query Match 54.7%; Score 1409; DB 10; Length 436;  
Best Local Similarity 59.6%; Pred. No. 2.5e-92;  
Matches 272; Conservative 56; Mismatches 89; Indels 39; Gaps 4;

QY 46 ARLSLVACAAAPAE-----APLSHVQALAEALAKPKDDPRKHVCYQVAAVRY 97  
DB 11 AGARHVAARAAAPTECCOCPPTPOAKLPHMQOALDELAKPKE--SRRLIMAOIAAVRY 68  
QY 98 AIAETGLGATTPGATTPGALGRLGFEDEVFTLFGADLTMEESSELLHRLTEHFAH 157  
DB 69 AIAETGLGATTPGATTPGALGRLGFEDEVFTLFGADLTMEESSELLHRLTEHFAH 128  
QY 158 HSDEPLPMTSCCPGNIAMLENSYDPLIFVSSCKSPOMLAAVKSYLEAKGIAPKDM 217  
DB 129 NKEEPLPMTSCCPGNIAMLENSYDPLIFVSSCKSPOMLAAVKSYLEAKGIAPKDM 188  
QY 218 VAVSIMPCRKQSEADRDPCVDADPTLQDLHVTITVELGNIFKRGINLAELPGEND 277  
DB 189 CNVSVAPCVKQGEADREWFNTGAGLADVDHVTAAVGTAEVKIFLRGKILNLPSPND 248  
QY 278 NPMGVSGAGVLTGTGVMALRTAYELFTGTPRLSLSEVRGMDGKIKETNITMVA 337  
DB 249 NPIGSETGALLFTGTGVMALRTAYELFTGTPRLSLSEVRGMDGKIKETNITMVA 308  
QY 338 PSKPEELLKRRARAEEAANGTGPPLAMDGAGFTSEDRGCTLRVAANGILGNANK 397  
DB 309 DDSPEK-----AFAGADGQ--GITLKIAVANGILGNANK 339  
QY 398 LITKMOAGEAKYDFEIMACPGVGGGPPSTDAITOKQAALYNDEKTLKRSHE 457  
DB 340 LTKSISEGAKYDFEIMACPGVGGGPPSTDAITOKQAALYNDEKTLKRSHE 399  
QY 458 NPSIRELYDTYLGEPLGRHAHELHHTYVAGVEEK 493  
DB 400 NPIQALYDKPLGAPSRHKAHDLHTHYVAGGIPEE 435

RESULT 4  
Q9AR66 PRELIMINARY; PRT; 449 AA.

AC 09AR66: 01-JUN-2001 (TREMUREL. 17, Created)  
DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)  
DE 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
DE Fe-hydrogenase precursor.  
GN HYDA.  
OS Scenedesmus obliquus.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales;  
OC Scenedesmeaceae; Scenedesmus.  
OX NCBI\_TaxID=3088;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21125815; PubMed-11096090;  
RA Florin L., Tsokogiou A., Happe T.;  
RT A novel type of iron hydrogenase in the green alga Scenedesmus  
obliquus is linked to the photosynthetic electron transport chain.;  
RL J. Biol. Chem. 276:6125-6132(2001).  
DR EMBL: AJ271546; CAC34419.1; -.  
DR HSSP: P29166; 1FEH.  
DR InterPro: IPR004108; Fe\_hyd\_1g\_C.  
DR InterPro: IPR003149; Fe\_hyd\_SSU.  
DR Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
KW Transit peptide.  
FT TRANSIT 1 35 POTENTIAL.  
SQ SEQUENCE 449 AA; 48540 MW; 9ABFC5E69580B458 CRC64;

Query Match 53.7%; Score 1384.5; DB 10; Length 449;

Best Local Similarity 58.2%; Pred. No. 1.4e-90;  
Matches 285; Conservative 52; Mismatches 98; Indels 55; Gaps 9;

QY 11 AVSIRSSCARQVAPAPLAAS---TVAVALATEAPARRILGNVACAAAPAPAPUS 66  
DB 11 AVSVR-----PPNVRAVVAADERRRLVVAAPTAECDD-----CPPAPAPAP-- 53  
QY 67 HVOQALAEALAKPKDDPRKHVCYQVAPAVVAALAEPLGALGATTPGOLAEGRGFE 126  
DB 54 HMOQTLDELAKPKEQ--RKVMAQIAPAVVAALAEPLGALGATTPGOLAEGRGFE 111  
QY 127 VFDTLFGADLTMEESSELLHRLTEHFAHSHDEPLPMTSCCPGMIAMLEKSYDPL 186  
DB 112 VFDTLFGADLTMEESSELLHRLTEHFAHSHDEPLPMTSCCPGMIAMLEKSYDPL 171  
QY 187 YVSSCKSPOMLAAVKSYLEAKGIAPKDMVNSIMPCRQSEADRDPCVDADPTLR 246  
DB 172 YLSCKSPOMLAAVKSYLEAKGIAPKDMVNSIMPCRQSEADRDPCVDADPTLR 230  
QY 247 QLDHVTITVELGNIFKRGINLAELPGENDNPMGVSGAGVLTGTGVMALRTAYE 306  
DB 231 NVDHVTITVELGNIFKRGINLAELPGENDNPMGVSGAGVLTGTGVMALRTAYE 290  
QY 307 LFTGTPLPRLSLSEVRGMDGKIKETNITMVAPEKPEELLKRRARAEEAANGTGPPLA 366  
DB 291 VVTKPLDRIVFEDVGLGKISTJLHLPGPSPE-----KAFAGADGT----- 335  
QY 367 WDGGAGFTSEDRGCTLRVAANGILGNANKLITKMOAGEAKYDFEIMACPGVGGG 426  
DB 336 -----GITLNIIVANGILGNANKLITKMOAGEAKYDFEIMACPGVGGG 381  
QY 427 QPRSTDAITOKQAALYNDEKTLRSHENPSIRELYDTYLGEPLGRHAHELHHTYV 486  
DB 382 QPRSAQKQILQKQAMTDLDERAVIRSHENPLIGALYKELGEPGRHAHELHHTYV 441  
QY 487 AGVEEKDEX 496  
DB 442 AGGV--PDEK 449

RESULT 5  
Q9AU60

ID 09AU60 PRELIMINARY; PRT; 403 AA.

AC 09AU60: 01-JUN-2001 (TREMUREL. 17, Created)  
DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)  
DE 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
DE Fe-hydrogenase (Fragment).  
OS Scenedesmus obliquus.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales;  
OC Scenedesmeaceae; Scenedesmus.  
OX NCBI\_TaxID=3088;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX SYRAIN-WILD TYPE D3;  
RA Munschiers R., Stangier K., Senger H., Schulz R.;  
RT Molecular evidence for a Fe-hydrogenase in the green alga Scenedesmus  
obliquus.;  
RL Curr. Microbiol. 42:353-360(2001).  
DR EMBL: AF276706; AAG59621.1; -.  
DR HSSP: P29166; 1FEH.  
DR InterPro: IPR004108; Fe\_hyd\_1g\_C.  
DR InterPro: IPR003149; Fe\_hyd\_SSU.  
DR Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
FT NON\_TER 1 1  
SQ SEQUENCE 403 AA; 43643 MW; 46110F53C137DA7F CRC64;

Query Match 46.4%; Score 1196; DB 10; Length 403;  
Best Local Similarity 57.7%; Pred. No. 3.3e-77;  
Matches 248; Conservative 48; Mismatches 96; Indels 38; Gaps 9;

OY		67	HVQOALAEIAKAKDPDTRKHVCQVAPAVRAVAIAEFLGLAIPGATTTKQAIEGRIRIGFDE	126
Db		2	HHOOITLDELAKKE--RKVMIAQLAPAVR-GIAETWGINPGDVTVGMONTGIRMIGFDY	57
OY		127	VFDTLFGADLTIMEGSELHRLTTEHLNAHPHSDEPLPMTSCCPGMIAMLEKSYDLPJ	186
Db		58	VFDTLFGADLTIMEGTELHRLDLQHNPKEERLPMPTSCCPGVANVEKSNELIP	117
OY		187	YVSSCKSPQMMAAAVKSLAEKKGIAPRDQMVVSIMPCTRKOSEADRDFCVADPTLR	246
Db		118	YLSSCKSPQMMAAIKNFAAEAGAKRPEDICNVSPVCVRKSGEAEPRSGSTHNHAGR	177
OY		247	QLDHITTYELCINIKERCINILAEIPEGEDMDPMVGSGAGVLFTGTGGMEALRTAYE	306
Db		178	DVDHMTTAELCKIVERKIKNELIOESFPDVPVCGSGGLFTGTGGMEALRTIVE	237
OY		307	LFTGPL--PRLSLSEVRGMDIKETNITMVPAPSGFEELLKRNARAABAAAHTPG-	363
Db		238	VVTAAEPQRGSRL-----TSTAMPFA-----QRSPPSRPQAPTAPSR	277
OY		364	PLANDGGAGFTSDDRGGITLLEVANAANGINKKLITKAQAGEAKDYEWELMACPAQYG	423
Db		278	PL-----QAOTGESGITLINIAVANNGINKKLITKOLAAGESKYDEFEMVACPGGCIG	328
OY		424	GGGCRSTDKAITKOROAALYNLDEKSTRSHENPSITELYDTYLGEPFLGHANHELHT	483
Db		329	GGGCRQ-RNKQLQKRQAAMDLDERRAVTRT-ENPLIALITEKFLGEEPNGRAHELLHT	386
OY		484	HYVAGVEEK 493	
Db		387	HYVAGVDPDR 396	
 RESULT 6 O9WY44 ID PRELIMINARY: PRT: 608 AA. O9WY44:				
AC		01-NOV-1999	(TREMBLrel. 12, Created)	
DT		01-NOV-1999	(TREMBLrel. 12, Last sequence update)	
DT		01-MAR-2002	(TREMBLrel. 20, Last annotation update)	
DE		MADP-reducing hydrogenase, subunit D, putative.		
CN		TMO201.		
OS		Thermotoga maritima.		
OC		Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.		
OX		NCB1_Taxid=2336;		
RN		[1]		
RP		SEQUENCE FROM N.A.		
KC		STRAIN=MSB8 / DSM 3109;		
KX		MEDLINE=99287316; PubMed=10360571;		
RA		Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,		
RA		Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,		
RA		McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,		
RA		Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson D.,		
RA		Heidelberg J., Sutton G.G., Fleischmann R.D., Olsen J.A., White O.,		
SA		Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;		
RT		"Evidence for lateral gene transfer between Archaea and Bacteria from		
RL		genome sequence of Thermotoga maritima."		
DR		Nature 399:323-329(1999).		
DR		EMBL; AF001705; AAC35293.1; -.		
DR		HSSP; P29166; IFSH.		
DR		TIGR; TMO201; -.		
DR		InterPro; IPR001450; 4Fe4S_ferredoxin.		
DR		InterPro; IPR001041; Ferredoxin.		
DR		InterPro; IPR004108; Fe_hyd_1g_C.		
DR		InterPro; IPR003149; Fe_hyd_SSU.		
DR		Pfam; PF00111; fer2; 1.		
DR		Pfam; PF00037; fer4; 2.		
DR		Pfam; PF02906; Fe_hyd_1g_C; 1.		
DR		Pfam; PF02256; Fe_hyd_SSU; 1.		
DR		PROSITE; PS00198; 4Fe4S_FERREDOXIN; 1.		
KX		Iron-sulfur. Complete proteome.		
CO		SEQUENCE 608 AA: 67652 MW; BC90822E35F905A6 CRC64;		

Query Match	39.2%	Score 1010.5	DB 16	Length 608
Best Local Similarity	47.9%	Pred. No. 1e-63		
Matches	216	Conservative	56	Mismatches 112; Indels 67; Gaps 8
QY	49	IGNVCAAAAP--AAEAPLSHWQOALAEIAKPPADDEPTRKHVCQVAPAVVAIAETLGLAP	107	
DB	218	ICGGCAACAFPCPTGLAVENSAYKVLELEKRE-----KLVVQTAPSRVALGEEFGYAP	272	
QY	108	GATTTCOLAEGLRGLGDFEVPFTLGLGADLTTEEGSELHRLTEHLGAHPSD-EPLPYF	166	
DB	273	GTISGQWAAALRRIGFDFEPTNGADLTTEEGSEPLEK-----GDLEDLPWF	325	
QY	167	TSCCGWIMLEKSYDPDLPYVSSCKSPQMMAAVYASYLEKKGIAPKDMVMSIMPT	226	
DB	326	TSCCGWVNLVKKYPERLTRLSAKSPQGMASAVKTYPAKLGKVEDIPHSIMCT	385	
QY	227	KROSEADRDWFCVADPTLROLDHVITTVELGNIKEKGINIAELPEGEWMPMGVSGA	286	
DB	386	AKKDEALROLVNVGVPV---VDVVLTRRELKLTIRMKKIPANPEEYDAPLGISTGA	442	
QY	287	GVLPECTGGVMEALRTAVELFTGCPRLSLSEVRGDKIKENTITVAPAGSFEBLL	346	
DB	443	AALEFVTGVMMAALRTAIELTKGALPPIVEEVRGLKNGREAIIDL-----	490	
QY	347	KHRAAARAEAAHCTPGPLAMDGAGFTSEDRGCTLRVAVANGLAGNAKLTMKQAGT	406	
DB	491	-----DCK---KIRIAVNGSTAVRNVLVEKILRE	517	
QY	407	AKYDVELIMACPGVGGGGGPRSDTKAIYQKRAALYNLDEKSLRSHSEPSIRELYD	466	
DB	518	VKYNHEVWACGGGIGGGGQPYSDPEILRRAAIYITIDERMTLRKSHENPAIKKLYE	577	
QY	467	TYLGEPLGKHAHELHHTHYVAGVGEKEDEK 497		
DB	578	EYLEHPLSHKAHELHHTY-----EDRRKK 603		
RESULT 7				
Q8RBC8	PRELIMINARY:	PRT:	581 AA.	
AC	Q8RBC8:			
DT	01-JUN-2002 (TREMBlrel. 21, Created)			
DT	01-JUN-2002 (TREMBlrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)			
DE	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kDa subunit (chain 6).			
GN	NDUG OR TTE0894.			
OS	Thermoaerobacter tengcongensis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;			
OC	Thermoaerobacteriales; Thermoaerobacteriaceae; Thermoaerobacter.			
NCBI	_TaxID=119072;			
RA	SEQUENCE FROM N.A.			
RP	STRAIN-MB4T / JCM11007;			
RA	MEDLINE=21992816; PubMed=11997336;			
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,			
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,			
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;			
RT	A complete sequence of T. tengcongensis genome."			
RL	Genome Res. 12:689-700(2002).			
DR	EMBL; AE013056; AAM24150.1; -			
KM	UBIquione; Complete proteome.			
SO	SEQUENCE 581 AA; 6446 MW; BC95C3203A9B1EB7 CRC64;			
Query Match	36.3%	Score 935.5	DB 16	Length 581
Best Local Similarity	43.2%	Pred. No. 2.1e-58		
Matches	189	Conservative	71	Mismatches 110; Indels 67; Gaps 7
QY	52	VACAAAPAAAPLSHWQOALAEIAKPPDDPTRKHVCQVAPAVVAIAETLGLAPGTT	111	
DB	199	AVCPYGAATYEDHTRKRYEALA-----DD--KRTVAAQAPAAVVAIAEGEFGAVGTV	250	
QY	112	PKOLAEGRLRLGFEDEVFTLEGADLTTEEGSELHRLTEHLGAHPSDDEPLPFTSCCP	171	

```

Db      251  TGNMAALRMGFAVFDNFPAADLTMEEGSELLERI-----KHGK-LPMITSCSP 302
QY      172  GWIMALEKSYDPLIPYSSCKSPOMMLAAMKSYIAEKKGIAPKDMVAVSIMPCTRKQSE 231
Db      303  GWIMCEKXYEFDIDNLSTCKSPHMMGALVKSYYAEKKGDIDPKDIPFVSIIMPCRAKLE 362
QY      232  ADROMFCVADPTLRDLHVTITVELGNIFKRGINLDELPEGENDNPMGVSAGVLEFG 291
Db      363  IEREMIRNG---MKDVAVLITRLRLAMIKEMGIDFVNLKDEEDDELGNSTGGAITFG 419
QY      292  TTGGVMEALRTAVELFTGTPPLRLSLSEVRGMGDIKETNITMVPAPSKFEELLKHRA 351
Db      420  ATGCVMEALRTVAIEVGRDGIKIDFEEVRGLGEGVRBATITI-----462
QY      352  ARAEAAHGTGPIPLAMDAGAGTSEDGCGITLRVAVANGIGNAKLITKMOAGEAKYDF 411
Db      463  -----DGMIDKIAIANGTGNARKLLDKYVAGEVEYHF 494
QY      412  VEIMACPAVGVGGGQOP---RSTOKAITOKROALYNLDEKSTLRSHENPSIRELYDTY 468
Db      495  IEVMCCPGGCGIMGGQPIHNPMEVEVKLRKAKAIYEIDKNLPIKSHENPAIKRLYEEF 554
QY      469  LGEPLGKRAHELHATHY 485
Db      555  LGYPLSEKSHELLATHY 571

```

## RESULT 8

Q9XC55 PRELIMINARY: PRT; 579 AA.

```

ID      Q9XC55
AC      Q9XC55;
DT      01-NOV-1999 (Tremblrel. 12, Created)
DT      01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT      01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE      Hydrogenase-1 (EC 1.18.99.1) (Fragment).
GN      HYDA.
OS      Clostridium thermocellum.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC      Clostridiales; Clostridiaceae; Clostridium.
OX      NCBI_Taxid=1515;
RN      [1]
RA      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 27405;
RA      Desai S.G., Stevens D.R., Stevenson D.M., Prince H.L., Gueriot M.L.,
RA      Lynd L.H.;
RT      Direct Submission.
RT      Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF148212; AAD3071.1; -.
DR      HSSP; P00195; 1CLEF.
DR      InterPro; IPR001450; 4Fe4S_ferredoxin.
DR      InterPro; IPR001041; Ferredoxin.
DR      InterPro; IPR004108; Fe_hyd_19_C.
DR      InterPro; IPR003149; Fe_hyd_SSO.
DR      Pfam; PF00111; fer2; 1.
DR      Pfam; PF00037; fer4; 2.
DR      Pfam; PF02906; Fe_hyd_1a_C; 1.
DR      Pfam; PF02256; Fe_hyd_SSO; 1.
DR      PRINTS; PR00353; 4Fe4SFERROXIN.
DR      PROSITE; PS00198; 4Fe4S_FERREDOXIN; 1.
KW      Iron-sulfur; Oxidoreductase.
FT      NON_TER
FT      SEQUENCE 579 AA; 63584 MW; 4D5A5ED4FA526DC7 CRC64;

```

## Query Match

Best Local Similarity 41.1%; Score 903.5; DB 2; Length 579;

Matches 198; Conservative 61; Mismatches 146; Indels 77; Gaps 7;

```

QY      17  SSCRAQVAPRAPPLASTYRVALATL-----EAPARRLGNV--ACAAAPAAEADPLSHY 68
Db      154  SACKNVQTVGAIDVTERGFRFTVSTAFNMPLSEVPVNCGCCINVCVPGALREKDDIDYV 213
QY      69  QQALAEIAKPKDDPFRKHVCQVAVPAVVAIAETIGLAGAATTPKQLAGLRLRGDFEYF 128

```

```

Db      214  WEALAN-----DELHVAVQFAPAVRALGEEFGMIGSRVTKMVAALSRLGFKKYF 265
QY      129  DTLFGADLTMEGSELLRLRLEHLEAPHSDPELPMTSCCPGMIAMLEKSYDPLIYV 188
Db      266  DDTFADLTMEGSELLNRI-----KNCGKLIITSCSPGMIKCEHNPPEFLNL 317
QY      189  SSCSPOMMLAAMVAVSYIAEKKGIAPKDMVAVSIMPCTRKQSEADRDNFCVADPTLRQL 248
Db      318  SSCSPHEMFQVAVLSYIAQKNGIDPSKVFVOSIMPCTRAKKEADRPBELSGTGP---DY 374
QY      249  DHVITVELGNIFKRGINLAEIPGENDNPMGVSAGVLEFGTTGVMEALRTAVELF 308
Db      375  DVVLITRELARIRKETGIDFNSLPDKQFDDPMGASGAGVYIGATGVMEALIRVGEILL 434
QY      309  TGTPPLRLSLSEVRGMGDIKETNITMVPAPSKFEELLKHRAAAAEAAAHOTPPRLAMD 368
Db      435  SKRPADKILEYTERVRLDGIKEASIEL-----460
QY      369  GGAGFTSEDRGCGITLRVAVANGIGNAKLITKMOAGEAKYDFVEIMACPAVGVGGQOP 428
Db      461  -----DGFLLKAAVANGIGNARKLIDKIRAGEADYFIETIMACPGGCGINGGQOP 509
QY      429  RSTD-----KAITOKROALYNLDEKSTLRSHENPSIRELYDTYLGEPHGKRAHELLHT 483
Db      510  IOPSSVYRWKMDIRCEERAKAIYEDESLPIRKSHENPKIMLYEEFFGEPGSHKRAHELLHT 569
QY      484  HY 485
Db      570  HY 571

```

## RESULT 9

Q8TG63 PRELIMINARY: PRT; 555 AA.

```

ID      Q8TG63
AC      Q8TG63;
DT      01-JUN-2002 (Tremblrel. 21, Created)
DT      01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT      01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE      [Fe]-hydrogenase (Fragment).
OS      Ptilomyces sp. E2.
OC      Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
OC      Neocallimastixaceae; Ptilomyces.
OX      NCBI_Taxid=73868;
RN      [1]
RA      SEQUENCE FROM N.A.
RC      STRAIN-E2;
RA      MEDLINE-21688625; PubMed-11891051;
RA      Voncken F.G.J., Boxma B., van Hoek A.H.A.M., Akhmanova A.S.,
RA      Vogels G.D., Huyen M., Veenhuis M., Hackstein J.H.P.;
RT      "A hydrogenosomal [Fe]-hydrogenase from the anaerobic chytrid
RT      Neocallimastix sp. L2."
RL      Gene 284:103-112(2002).
DR      EMBL; AF446076; AAL90459.1; -.
FT      NON_TER
FT      SEQUENCE 555 AA; 62033 MW; AEF0134FBB63B18 CRC64;

```

## Query Match

Best Local Similarity 34.7%; Score 893; DB 3; Length 555;

Matches 181; Conservative 67; Mismatches 105; Indels 54; Gaps 7;

```

QY      81  DPTFRKHVCQVAVPAVVAIAETIGLAGAATTPKQLAGLRLGDFEYFDTLFGADLTME 140
Db      186  DTRKRTVVSTAPATVAVLAEEFNAPDDDTFGKMAVAGLKLGFDTLIFDTRNSADLTME 245
QY      141  EGSELLRLRLEHLEAPHSDPELPMTSCCPGMIAMLEKSYDPLIYVSSCKSPOMMLAA 200
Db      246  EGSELLTRLNE-----GKRPMTSCCPGMINNVEKSYPIRBNLSSCKSPQOMIGA 297
QY      201  MVSYSIAEKKGIAPKDMVAVSIMPCTRKQSEADRDNFCVADPTLRDLHVTITVELGNI 260
Db      298  VIKTFPAKKINKPEDIIVHSVMPCTAKKGAEKRPFRKRDGPV---DIDHVITTRRLITL 354

```

QY 261 FKRGINLAELPEGENDNPGVSGAGVLTGTGVMALRTAYELFTG--DTPLRLSL 318  
 Db 355 LKKRINPSELKNEKEDSPGLIGSSAGNLEFGVYTGVMALRTA-QIITGVENPIELGEL 413  
 QY 319 SEVRGMDGKETNITVPAVPGSKFEELKHRAAARAAAHGTPGLANDGAGFTSEDO 378  
 Db 414 KAIRGIDGIRKASV-----PL-----KTKDG 434  
 QY 379 RGITLRVAVANGSKALLITKMOAGEAKYDFEIMACPGAGGGGPRSDKAITOK 438  
 Db 435 K-DVNVRAAVVSGANIQFLKLEKKELEPFDEVMKCGGCGINGGOKPSADPRVAK 493  
 QY 439 ROALYNLDEKSTLRSHENPSIRELYDTYLGEPILKHAHELHHTHY 485  
 Db 494 KMERVMTMDQASRLSRSHENPEITQYKEFLKPNHLSHELHHTHY 540

RESULT 10  
 Q8TFP2 PRELIMINARY; PRT; 636 AA.  
 AC 08TFP2: 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hydrogenase.  
 GN HYD12.  
 OS Neocallimastix frontalis (Rumen fungus).  
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;  
 OC Neocallimastixaceae; Neocallimastix.  
 OX NCBI\_TaxID=4757;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21888625; PubMed=11891051;  
 RA Voncken F.G.J., Boxma B., Van Hoek A.H.A.M., Akhmanova A.S.,  
 RA Vogels G.D., Huynen M., Veenhuis M., Hackstein J.H.P.,  
 RT "A hydroxymethyl [Fe]-hydrogenase from the anaerobic chytrid  
 Neocallimastix sp. L2."  
 RL Gene 284:103-112(2002).  
 DR EMBL; AY033895; AAK60409.1;  
 SQ SEQUENCE 636 AA; 70449 MW; 4A4E78235FA3A0D7 CRC64;

Query Match 34.2%; Score 880.5; DB 3; Length 636;  
 Best Local Similarity 43.0%; Pred. No. 2e-54;  
 Matches 108; Conservative 67; Mismatches 123; Indels 59; Gaps 9;  
 QY 54 CAAAP-AAEAPLSHYQALAEELAKPKDPTKRHCVCYAPAVVAIAETGLAPGATTP 112  
 Db 238 CQGVCPVGAITARTEVVDLRLH-----DTKRKYVCSTAPAIRVAPAEFSEADPDT 292  
 QY 113 KOLAEGLRLGDEVEFDLTLEGADLTMEGSELLRLHLEHAPHSDEPLPFTSCCPG 172  
 Db 293 GKXVAALRLKGFYIDTNSADLTMEGTELDLRL-----NNGCKPFTSCCPG 344  
 QY 173 WTAMLEKSYPDILPYVSSCKSPOMMLAAVKSYLAKKGIAPKDMVMSIMPTKQSEA 232  
 Db 345 WTNWKSYPBLSNDLSCKSPOMMLGAVIKSTFAKLGISTEDIIHVSIMPTAKGEA 404  
 QY 233 DRDMFCVDA--DPTLRQLDHVITVYELGNIIFKRGINLAELPEGENDNPGVSGAGVLE 290  
 Db 405 RRPEFYQKRGKDDYDIDYITTRRELLLLKAKKINPAELPDKDSPLGIGSSAGNLF 464  
 QY 291 GTTGGMALARTAYELFTG--TPRLSLSEVRGMDGKETNITVPAVPGSKFEELKH 348  
 Db 465 GYVGCGVMAAIRTA-QVITGVENPIELGELKAIKRGIDGIRKASV-----507  
 QY 349 RAAARAAAHGTPGLANDGAGFTSEDRGGITLRVAVANGSKAKLITKMOAGEAK 408  
 Db 508 -----PL-----KTKDG--EVSRAAVVSGANIQFLKLEKKELE 543  
 QY 409 YDFEIMACPGAGGGGPRSDKAITOKROALYNLDEKSTLRSHENPSIRELYDTY 468  
 Db 544 PFIEEMKCGGCGINGGOKPSADPRVAKKMRVMTMDQASRLSRSHENPEITQYKEFL 603

QY 469 LGEPILKHAHELHHTHY 485  
 Db 604 LGEPNSHLAHELHHTHY 620

RESULT 11  
 Q46508 PRELIMINARY; PRT; 585 AA.  
 AC 046508: 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Potential NAD-reducing hydrogenase subunit.  
 OS Desulfovibrio fructosovorans.  
 OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;  
 OC Desulfovibrio.  
 OX NCBI\_TaxID=878;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95270577; PubMed=7751270;  
 RA Malki S., Saimmaime I., De Luca G., Rousset M., Dermoun Z.,  
 RA Belatch J.P.,  
 RT "Characterization of an operon encoding an NADP-reducing hydrogenase  
 in Desulfovibrio fructosovorans."  
 RL J. Bacteriol. 177:2628-2636(1995).  
 DR EMBL; U07229; AAA87057.1;  
 DR HSSP; P29166; 1FEH.  
 DR Interpro: IPR001450; 4Fe4s\_ferredoxin.  
 DR Interpro: IPR001041; Ferredoxin.  
 DR Interpro: IPR004108; Fe\_hyd\_19\_C.  
 DR Interpro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF00111; fer2; 1.  
 DR Pfam: PF00937; fer4; 1.  
 DR Pfam: PF02906; Fe\_hyd\_19\_C; 1.  
 DR Pfam: PF02286; Fe\_hyd\_SSU; 1.  
 DR PRINTS; PR00353; 4FE4SFRODOXIN.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1.  
 KW Iron-sulfur.  
 SQ SEQUENCE 585 AA; 63429 MW; C14D0E6FA3E2A72 CRC64;

Query Match 34.0%; Score 877; DB 2; Length 585;  
 Best Local Similarity 41.4%; Pred. No. 3.1e-54;  
 Matches 192; Conservative 66; Mismatches 138; Indels 68; Gaps 9;  
 QY 33 STVRVALATLEAPARRLGNV-----CAAAP-AAEAPLSHYQALAEELAKPKDPT 82  
 Db 173 SCVNRGFTVAVNAPEFM-NADVYTCNCGCVAVCPGALVEHEITVEYEAALND--- 228  
 QY 83 TRKHVCYAPAVVAIAETGLAPGATTPKOLAEGIRLGEVEFDLTLEGADLTMEEG 142  
 Db 229 --KVIVIQAPAAVRAALGIEDLVAPGTSVTKMAALRLRGLFPHVFDTPAALDTIMEEG 286  
 QY 143 SELHLRTELEHNPISDEPLPFTSCCPGIMLEKSPDILPYVSSCKSPOMMLAAV 202  
 Db 287 SEFLRLGKHLAC--DTNKLPLTLTSCCPGVAFFEHQEPDMADVPSTASPOOMGALA 344  
 QY 203 KSYLAEEKGIAPKDMVMSIMPTKQSEADRDMPFCVDA-DPTLRQLDHVITVYELGNIIFK 262  
 Db 345 KTYVALDLGIPRKLVLVVSMPCLAKKYBCARPEFVGNP---DVDIYITTELKLVK 401  
 QY 263 ERGINLAELPEGENDNPGVSGAGVLTGTGVMALRTAYELFTGTPPLPLSEVR 322  
 Db 402 RMNIDEPAGDEDEDPALGASTGAAPIFGTGIVIEAALRTAYELATGETLKGDEVEDY 461  
 QY 323 GMDGKETNITVPAVPGSKFEELKHRAAARAAAHGTPGLAMPGAGFTSEDRGGI 382  
 Db 462 GMDGVAKAKV-----GDN 476  
 QY 383 TLRAVANGSKAKLITKMOAGEAKYDFEIMACPGAGGGGPR--RSTDKAITOKROA 441  
 Db 477 ELVIGVAHNGNRELKRCOGAGET--FHAIEVACPGGCGGCGGYHHGDVLELKKRTO 535  
 QY 442 ALYNLDEKSTLRSHENPSIRELYDTYLGEPILKHAHELHHTHY 485



```

Db 462 -----
QY 399 IRKMGAEKAYDFVEIMACPVGGGPP-----ASTDKA-ITOKQAALYNLDEKSTLR 453
Db 481 MOLVRGKADHYFIETIMGSSGGCVGGPPHVSFKREKYNVKLERAKALYEDKLDRK 540
QY 454 RSHENPSIRELYDTYLGEPLGKHAHELHHTHY 485
Db 541 KSHHNSVAKRYEYLGRKNGKHAHELHHTHY 572

RESULT 14
Q9GTP2 PRELIMINARY: PRT: 589 AA.
AC 09GTP2:
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Putative 64kDa iron hydrogenase (Fragment).
OS Trichomonas vaginalis.
OC Eukaryota; Parabasallidae; Trichomonadidae; Trichomonadidae;
OC Trichomonas
ON NCBI_TaxID=5722;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20523972; Pubmed-11070057;
RA Horner D.S., Foster P.G., Embley M.T.;
RT "Iron Hydrogenases and the Evolution of Anaerobic Eukaryotes.";
RL Mol. Biol. Evol. 17:1695-1709(2000).
DR EMBL: AF262401; AAG31037.1; -.
DR HSSP: P29166; IFEH.
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR InterPro: IPR004108; Fe_hyd_19_C.
DR InterPro: IPR003149; Fe_hyd_SSU.
DR Pfam: PF00111; fer2; 1.
DR Pfam: PF00037; fer4; 2.
DR Pfam: PF02906; Fe_hyd_19_C; 1.
DR Pfam: PF02256; Fe_hyd_SSU; 1.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 1.
FT NON_TER
SQ SEQUENCE 589 AA; 64854 MW; F0B69ED78B9D066A CRC64;

Query Match 31.3%; Score 805.5; DB 5; Length 589;
Best Local Similarity 36.2%; Pred. No. 4e-49;
Matches 177; Conservative 79; Mismatches 132; Indels 101; Gaps 8;

QY 6 LKPCAAVSTRGSSCRARQVAPRAPLASTYRVALATLEAPARRLGNAVACAAAPAAEAPL 65
Db 187 IQPTFVTLQETSC-----IKCGCTLYCPV-----GAITEK 218

QY 66 SHVOQALAEIAKDKDPTKRHCVOYAPAVRAIAETLGLAPGATTPKQALBGLRLGPD 125
Db 219 SQVKEADLITA-----NKGKTIYQVAPRAVALSEAFGKBEVTGKAVSALKALGPD 274

QY 126 EVDFTFGADLTIMEGSELHRLTEHLEAHPHSDEPLPFTSCCGIMALEKSPDLI 185
Db 275 LVEDTNKADLTICEBAGELVNLRD-----PNAKFPHFTCCCAWVYVQSAPDFI 327

QY 186 PIVSSCKSPQGMILAAVKSYLEAKGCIAPKQWVNSIMPCTRKQSBADRMFCVADPTL 245
Db 328 PNLSSCRSPQGMILAKINYLPLKLDVQKEDVLFNSIMPCTAKKDEVER-----PEL 379

QY 246 R-----QLDHVTYVELGIFERGINLAELPEGEEDNPMGVSGAGVIFGTGVMFA 299
Db 380 RTSGKRETDVMTLVLELMIKLSINDEFNNLPDQDFNFGFSGAGQIFAAFGVMEFA 439

QY 300 ALTAATELFTGTPRLPRLTSEVGMGDIKFTNTITVVPAPGSKFEELLKRAAARAFAAH 359
Db 440 ASATAEYVYTGKLTNNITPVAGMGQIRAEIDL----- 474

QY 360 GTFGPLAMDGAGFTSEDRGGITLRVAVANGIGNAKKLTTRKMGAEAK---YDVEIMA 416

```

```

Db 475 -----
QY 417 CPAGCGGGGPPRSTDKATITOKQAALYNLDEKSTLRSHENPSIRELYDTYLGEPLGK 476
Db 515 CPGGCGGGTQPPKRNVSILNRNLAITYIDAKMECRKSHENPLINGVKEPLGKRNSHL 574
QY 477 AHELHHTHY 485
Db 575 AHELHHTHF 583

RESULT 15
Q46606 PRELIMINARY: PRT: 606 AA.
AC 046606:
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Hyd gamma.
OS Desulfovibrio vulgaris (strain Hildenborough).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
ON NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-HILDENBOROUGH;
RA Stokkermans J.P.M.G.;
RT Submitted (JAN-1991) to the EMBL/GenBank/DBJ databases.
RL
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-HILDENBOROUGH;
RA Stokkermans J., van Dongen W., Kaan A., Van Den Berg W., Veeger C.;
RT "Hyd Gamma, a gene from Desulfovibrio vulgaris (Hildenborough) encodes
a polypeptide homologous to the periplasmic hydrogenase.";
RL FEMS Microbiol. Lett. 58:217-222(1989).
DR EMBL: X57838; CAA40970.1; -.
DR HSSP: P29166; IFEH.
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR InterPro: IPR004108; Fe_hyd_19_C.
DR InterPro: IPR003149; Fe_hyd_SSU.
DR Pfam: PF00111; fer2; 1.
DR Pfam: PF00037; fer4; 2.
DR Pfam: PF02906; Fe_hyd_19_C; 1.
DR Pfam: PF02256; Fe_hyd_SSU; 1.
DR PRINTS: PR00353; 4FE4SFEROXIN.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 1.
FT Iron-sulfur.
SQ SEQUENCE 606 AA; 66030 MW; F3EB14EA04C347D6 CRC64;

Query Match 30.8%; Score 794.5; DB 2; Length 606;
Best Local Similarity 42.4%; Pred. No. 2.5e-48;
Matches 169; Conservative 61; Mismatches 112; Indels 57; Gaps 6;

QY 90 QVAPAVRAIAETLGLAPGATTPKQALBGLRFPDEVDFTFGADLTIMEGSELHRL 149
Db 233 QFAPAVRVGGEFGILPGSSVGVPTALRLGADVLDITNAADLVIMESTELLQRL 292

QY 150 TEHLEAHPHSDEPLPFTSCCGIMALEKSPDLIPYSSCKSPQGMILAAVKSYLEAK 209
Db 293 -----BGAKLPLFTSCCPGWVFAEKHLPLDIPVSTFRSPQCGALAKTYLART 344

QY 210 KGIAPKQWVNSIMPCTRKQSBADRMFCVADPTLRQDHYTTYVELGINFERGINLA 269
Db 345 MNYAPRRMRVNSIMPCTAKKEAARPEFRDG--VRDVAVLTTFEFARLRREGIDLA 401

QY 270 ELPEGEEDNPMGVSGAGVIFGTGVMFAALRTAYELFTGTPRLPRLTSEVGMGDIK 328
Db 402 GLEPSPCDPLMKRATGAIAVIFTGTGVMFAALRTYHYVUNGELAPVELHMLARGENVR 461

QY 329 ETNITVVPAPGSKFEELLKRAAARAFAAHGTGPLAMDGAGFTSEDRGGITLRVAV 368

```



Db	462	EAUV-----	PL-----	GEONGSYKVAV	478
OY	389	ANGCNAKKLITKMOAGEAKYDFVEIMACPRACVGGGQPRS--	TDKAITOKROALYNL		446
Db	479	VHGLAAROMVEAVLAGADHVEVVMACPGGCMDDGGQPRSKRAYNPNAOARRAALFSL			538
OY	447	DEKSTLRSHENPSIRELYDTYLGEPPLGKHAHELLHTRY			485
Db	539	DAENALROSHNNPLIGKYESFLGEPSCNLSHRLHTRY			577

Search completed: June 3, 2003, 16:09:07  
 Job time : 88 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:38:18 ; Search time 27 Seconds

(without alignments)  
541.600 Million cell updates/sec

Title: CAC80065

Perfect score: 2576

Sequence: 1 MSALYKPCAAVSIRGSSCR.....HELLHTHYAGVGEDEKK 497

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/prodata/1/1aa/6A.COMB.pep.\*  
3: /cgn2\_6/prodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/prodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/prodata/1/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125.5	4.9	748	4 US-09-066-046-4	Sequence 4, Appli
2	112	4.3	4472	2 US-08-804-227C-2	Sequence 2, Appli
3	109.5	4.3	857	4 US-08-887-534A-47	Sequence 47, Appli
4	105.5	4.1	3491	2 US-07-642-734C-2	Sequence 2, Appli
5	105.5	4.1	3491	3 US-08-439-009A-2	Sequence 2, Appli
6	100.5	3.9	3170	4 US-09-036-987A-4	Sequence 4, Appli
7	100.5	3.9	3170	4 US-09-370-700-4	Sequence 4, Appli
8	99.5	3.9	638	2 US-08-426-125-4	Sequence 4, Appli
9	99.5	3.9	638	2 US-08-455-355-4	Sequence 4, Appli
10	99.5	3.9	3724	2 US-08-804-227C-10	Sequence 10, Appli
11	99.5	3.9	3724	2 US-08-804-198-4	Sequence 4, Appli
12	99.5	3.9	4550	2 US-08-804-227C-8	Sequence 8, Appli
13	99.5	3.9	4550	2 US-08-804-198-2	Sequence 2, Appli
14	98.5	3.8	823	4 US-09-134-001C-4081	Sequence 4081, Ap
15	94.5	3.7	1996	2 US-08-804-227C-9	Sequence 9, Appli
16	94.5	3.7	1996	2 US-08-804-198-3	Sequence 3, Appli
17	94	3.6	537	3 US-08-920-610-2	Sequence 2, Appli
18	94	3.6	805	4 US-09-513-783A-178	Sequence 178, App
19	94	3.6	2035	1 US-08-046-585-5	Sequence 5, Appli
20	94	3.6	2035	1 US-08-393-703-5	Sequence 5, Appli
21	94	3.6	2035	5 PCT-US93-11721-5	Sequence 5, Appli
22	94	3.6	4545	2 US-08-804-227C-14	Sequence 14, Appli
23	93.5	3.6	3782	4 US-08-105-537-4	Sequence 4, Appli
24	93	3.6	4551	3 US-09-320-878-1	Sequence 1, Appli
25	93	3.6	4613	4 US-09-105-537-31	Sequence 31, Appli
26	93	3.6	11877	4 US-09-105-537-6	Sequence 6, Appli
27	92.5	3.6	514	4 US-09-385-028-5	Sequence 5, Appli

28	92.5	3.6	530	3 US-09-222-817-12	Sequence 12, Appli
29	92.5	3.6	530	3 US-09-222-817-14	Sequence 14, Appli
30	92.5	3.6	530	4 US-09-222-786-12	Sequence 12, Appli
31	92.5	3.6	530	4 US-09-222-786-14	Sequence 14, Appli
32	92.5	3.6	1833	4 US-08-621-944A-4	Sequence 4, Appli
33	92.5	3.6	1833	4 US-08-945-567D-4	Sequence 4, Appli
34	92.5	3.6	1992	4 US-08-621-944A-3	Sequence 3, Appli
35	92.5	3.6	1992	4 US-08-945-567D-3	Sequence 3, Appli
36	92.5	3.6	2048	4 US-09-268-347-48	Sequence 48, Appli
37	92.5	3.6	2123	4 US-08-968-685A-10	Sequence 10, Appli
38	92.5	3.6	3588	4 US-09-036-987A-6	Sequence 6, Appli
39	92.5	3.6	5588	4 US-09-370-700-6	Sequence 6, Appli
40	92	3.6	607	3 US-09-335-409-9	Sequence 9, Appli
41	92	3.6	607	4 US-09-568-102-9	Sequence 9, Appli
42	92	3.6	607	4 US-09-567-969-9	Sequence 9, Appli
43	92	3.6	607	4 US-09-568-480-9	Sequence 9, Appli
44	92	3.6	607	4 US-09-568-486-9	Sequence 9, Appli
45	92	3.6	607	4 US-09-568-472-9	Sequence 9, Appli

#### ALIGNMENTS

RESULT 1  
US-09-066-046-4  
Sequence 4, Application US/09066046A  
Patent No. 6204252  
GENERAL INFORMATION:  
APPLICANT: MURPHY, Cheryl  
STOREY, James  
BELTZ, Gerald A.  
COUGHLIN, Richard T.  
TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
ENRICHED AND METHODS OF USE  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: United States  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/09/066.046A  
FILING DATE: 24-Apr-1998  
CLASSIFICATION: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Superko, Colleen  
REGISTRATION NUMBER: 39, 850  
REFERENCE/DOCKET NUMBER: 106.941.155  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-066-046-4

Query Match 4.9%; Score 125.5; DB 4; Length 748;  
Best Local Similarity 23.2%; Pred. No. 0.00055;  
Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

QY 248 LDHVTTELVEICNIFKRGINLAELPEGEMDNPMCGVGAGVLEFGTTGCVMEALRTAYEL 307  
 Db 98 LEHITIT-EVIVSNEE-----ITPBEK-----KTLTAEALTSKGVKALIKNSADV 146  
 QY 308 FTGTPRLSL-----SEVRGMDGJKEKITITVPAPGS----- 340  
 Db 147 -NMSPEPAITLGIQRCFGQSKAIKRLKRVAGAHINTPTGSPPLAAVQAAANEASNL 205  
 QY 341 -----KFEELKHHRAABAAEAAGTP-----GPIADGGAGFTSEDRGCGIT 383  
 Db 206 KEANKIYNFLHRRADLSTEHCTPALHLATAGNRTAMLLDKCAPATORDAKRTA 265  
 QY 384 LRAVANGIGNAKKILITKMOAGEAKYDVEIMACPGACGGGQPRSTDAITQKQOAL 443  
 Db 266 LHIAANGDGKLYRMIAK-----KCPDSC-----QPLCSMDGDTALHE-AL 305  
 QY 444 Y--NLDEK-----STLRSHENPSTRELYDTYLGEPLGKHAHELH 482  
 Db 306 YSDNVEKECFKMLKESRKLHLSNSFFGDLNTPQEAANGDTLLH 349

## RESULT 2

US-08-804-227C-2  
 : Sequence 2, Application US/08804227C  
 : Patent No. 5876991  
 : GENERAL INFORMATION:  
 : APPLICANT: Dehoff, Bradley S.  
 : APPLICANT: Kuester, Stuart A.  
 : APPLICANT: Rostock, Paul R., Jr.  
 : APPLICANT: Sutton, Kimberly L.  
 : TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 : NUMBER OF SEQUENCES: 15  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: THOMAS G. PLANT 1501  
 : STREET: LILLY CORPORATE CENTER  
 : CITY: INDIANAPOLIS  
 : STATE: IN  
 : COUNTRY: USA  
 : ZIP: 46285  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM compatible  
 : OPERATING SYSTEM: MS-DOS  
 : SOFTWARE: ASCII(DOS) Text only  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/804,227C  
 : FILING DATE: February 21, 1997  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Plant, Thomas, G.  
 : REGISTRATION NUMBER: 35,784  
 : REFERENCE/DOCKET NUMBER: X-8231  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 317-276-2459  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 4472 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: unknown  
 : MOLECULE TYPE: protein  
 : US-08-804-227C-2

Query Match 4.38; Score 112; DB 2; Length 4472;

Best Local Similarity 23.48; Pred. No. 0.29; Mismatches 124; Indels 128; Gaps 24;

Matches 91; Conservative 46; Mismatches 124; Indels 128; Gaps 24;

QY 143 SELLHRLTLEHLEAH-----HSDPELPMFTSC--PGWIMALEKSYDIL-- 185  
 Db 1032 AQLHL--TELLDSCPSGVLAPOKSPFAQEPILAVVGMGRFPGVSPALMRLVYEGV 1089  
 QY 186 -----PV-----VSSCKSPOMLAAVKSYL-----AEKKIAPKDMYVS 221

Db 1090 DAVSPFEGDRQMGVEGLYDEPVGAG--KSYVRBGLFLHDAEFDAAFFGISPREAVAMD 1147  
 QY 222 IMPCTRKROSEADRMFCVD--ADPTLRQDHYVTTVELGIFKRGINLAELPEGEMDN 278  
 Db 1148 -----PQORLLLETSMELIERAGIDP--HSLHGSTGYAGVMPQETPRALAEAE----- 1196  
 QY 279 PMGVSGAGVLEFGTTGCVMEALRTAYELEFGTPLP-----RLSLSEVNG-- 323  
 Db 1197 -----GSDGYLLTGTSGSVSG--RVAYTLLEGPAVVDVACSSSLVALHLAVQALRGE 1250  
 QY 324 MDGKEKITITVPAPGSKFEBELKHHRAA--RAEAAHGTPLADGGAGF----- 373  
 Db 1251 CDNALAGVYVMAGP--MFVEFSRQRLADGRCKARADGADG--TAAEAGVVLVERLS 1308  
 QY 374 -----TSEDG--RGGITL-----RVAVANGIGNAKKILITKMOAGEAK 408  
 Db 1309 DARRLGHVLAIVCGSVNODGASNGILAPSGPQERY--IRQALGNRLTVADVVEAH 1367  
 QY 409 -----YDFVEIMACPGACVG--GGGCP 428  
 Db 1368 GTGRLGDPIEAQAL-LGTGGRDRGGRP 1395

## RESULT 3

US-08-887-534A-47  
 : Sequence 47, Application US/08887534A  
 : Patent No. 6455323  
 : GENERAL INFORMATION:  
 : APPLICANT: Holden, David W.  
 : TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
 : NUMBER OF SEQUENCES: 106  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 : STREET: 233 South Wacker Drive/6300 Sears Tower  
 : CITY: Chicago  
 : STATE: Illinois  
 : COUNTRY: United States of America  
 : ZIP: 60606  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/887,534A  
 : FILING DATE:  
 : CLASSIFICATION: 536  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Rin-Laures, Li-Hsien  
 : REGISTRATION NUMBER: 33,547  
 : REFERENCE/DOCKET NUMBER: 28341/33996  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (312) 474-6300  
 : TELEFAX: (312) 474-0448  
 : TELEX: (312) 474-6600  
 : INFORMATION FOR SEQ ID NO: 47:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 857 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-08-887-534A-47

Query Match 4.38; Score 109.5; DB 4; Length 857;

Best Local Similarity 21.28; Pred. No. 0.033; Mismatches 162; Indels 91; Gaps 19;

Matches 86; Conservative 67; Mismatches 162; Indels 91; Gaps 19;

QY 108 GATTPKQIAGLRR-LGFDEVFTLFGADLTINEGSELHRLTLEHLEAHPH--SDEPL 163  
 Db 312 GATTLDEYRIYIEKDALERFQKVFAYAESV-EDTIALIRGLKERELHHHQTTPAI 370  
 QY 164 PMFTSCPGMIAMLEKSY-----DLIPYSSCKSPON-----MLA 199

Db 371 VAAATLSHRYIA--DROLPKDAIDIDEAASSIRMOIDSKPEELDRRIITOLKLEQQA 428  
 QY 200 AMVKYLAEEKGIAPKDMVMSIMPTCKROSEADRDWFCVAD---PTLR-OLDHVIT 254  
 Db 429 LMKEDEASKRL---DMLNEELSDKEROYSELEEWKAKSLSTOTIKAELEQAKTA 485  
 QY 255 VE---LGNIFERGINLAELPEGEWDMNMGVSGAGVLEFQT---GGVMEALRTAY 305  
 Db 486 IQQARVGDGLARMSLELOKIKPELEKQL-----EATOLEGKTMRLRKRYVDALEIAEVL 540  
 QY 306 ELFTGTPLRLSLSEVRGMDGKIKETNITVWPAPGSKFEELKHRAAARA----- 356  
 Db 541 AMWTGIPVSRMSESE-----REKLRR-----EQELHRRVIGQNEAVDAVSNMIR 585  
 QY 357 -AAHGTPLPLANDGAGTSEDGRTITLRVAVANGLSNAKLLITKMOGE--AKYDVE 413  
 Db 586 NSRAGLADPNRPISGLFPGTGVGTCLCKALANMFSDAMVAFIDSEMEKHSYR 645  
 QY 414 IMACPAGCVG---GGGPRSTDKAITOKRQALYNLDEKSTLRSH 456  
 Db 646 LVGAPPGVYGEEGYLTEA-----VRRRPYSVILLDE---VEKAM 683  
 RESULT 4  
 US-07-642-734C-2  
 ; Sequence 2, Application US/07642734C  
 ; Patent No. 5824513  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Katz, L  
 ; APPLICANT: Donadio, S  
 ; APPLICANT: McAlpine, J B  
 ; TITLE OF INVENTION: Recombinant DNA Method for Producing  
 ; TITLE OF INVENTION: Erythromycin Analogs  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Edward H. Gorman  
 ; STREET: Abbott Laboratories D377/Ab6D-2 One Abbott  
 ; STREET: Park Rd  
 ; CITY: Abbott Park  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60064-3500  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/642,734C  
 ; FILING DATE: 17-JAN-91  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Danckers, Andreas M  
 ; REGISTRATION NUMBER: 32652  
 ; REFERENCE/DOCKET NUMBER: 4952.US.01  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 708-937-9396  
 ; TELEFAX: 708-938-2623  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3491 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-07-642-734C-2  
 Query Match 4.1%; Score 105.5; DB 2; Length 3491;  
 Best Local Similarity 21.6%; Pred. No. 0.93;  
 Matches 117; Conservative 60; Mismatches 213; Indels 151; Gaps 25;

QY 66 SHVOAL---AELAKRDPDTRKHVCQVAPARVAIAETLG----- 104  
 Db 457 --VSAALIGEVNIALIYHPITPRALAEALAEVQAORTTRARTEAABGEVAVVYAMAC 514  
 QY 105 -LAPGATPKO---LAELIRL-----GFDEVDTLFGADLTIMBGSSELRHLEH 152  
 Db 515 RLPGVSTPEEFWELLSEBRDAVAGLPTRGMD--LDSLFHPDPT--RSGT----- 561  
 QY 153 LEAHPHSDPELPMFTSCCGWIAMLEKSPDLIPVYSSCKSPQMLAAMVYSTLAEKGI 212  
 Db 562 --AHQRCGFLTEATFADPAFGM-----SPREALAVDPOQRIMLELSWEVLERAGI 611  
 QY 213 APKDMVMSIMPTCKROSEADRDWFCVADPTLRDLDHITVTEELNIFERGINIAELP 272  
 Db 612 PPTSL-----QASP-----TGVFGLIPQETGPLAE-- 638  
 QY 273 EGEWDMNMGVSGAGVLEFQTGGVMEALRTAYELETGTPLRLSLSEV----- 321  
 Db 639 -----GGEVGEYLMGTITTSV--ASGRIVTL--GLEGPAISVDTCSSSIVAVHIA 687  
 QY 322 ----RGMDIKET-NITVWPAPGSKFEELKHRAAARAALAAHGTPLPLANDGAGPFS 375  
 Db 688 COSLRGESSLAMAGGVTVMPGMLYDFSRMNSILAPDCRKAFTA-----GANGFCM 740  
 QY 376 EDGRGITLRVAVANGLSNAKLLITKMOGEAKYDVEI-MACPAGCVGGGOPRSTDKA 434  
 Db 741 ABGAGMLLE-RUSDARRNGHPVLAVRGTAVNSDASNGLSNP-----NGRAOVRTQQA 795  
 QY 435 ITOK--ROAALYNLDEKSTLRSHENPSIRELYDTV---LGEP--LGHRAHELLATHRYA 487  
 Db 796 LAESGIGRADIAVEAHGHTGLGDIETEARLFEATGRBQPLHIGSVKSNLGHQAAA 855  
 QY 488 G 488  
 Db 856 G 856  
 RESULT 5  
 US-08-439-009A-2  
 ; Sequence 2, Application US/08439009A  
 ; Patent No. 6004787  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Donadio, S  
 ; APPLICANT: Katz, L  
 ; APPLICANT: McAlpine, J B  
 ; TITLE OF INVENTION: Method of Directing Biosynthesis of  
 ; TITLE OF INVENTION: Specific Polyketides  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Steven F. Weinstock  
 ; STREET: Abbott Laboratories D377/Ab6D-2 One Abbott  
 ; STREET: Park Rd  
 ; CITY: Abbott Park  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60064-3500  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/439,009A  
 ; FILING DATE: 11-MAY-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Casuto, Dianne  
 ; REGISTRATION NUMBER: 40,943  
 ; REFERENCE/DOCKET NUMBER: 4952.US.D1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 847-938-3137  
 ; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 3491 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-439-009A-2

Query Match 4.1%; Score 105.5; DB 3; Length 3491;  
 Best Local Similarity 21.6%; Pred. No. 0.93;  
 Matches 117; Conservative 60; Mismatches 213; Indels 131; Gaps 25;

10 AAVSRGSSCARQAPAPAPLAATVVALATLEPA---RLGNVCAAAAPAAEAPL 65  
 DB ASTGRHRTCCERLLA---VVGNETAALAGREADAATRELGIDSVLAQLRAK--- 456  
 QY SHVQOAL---AELAPKDDPTTRKHVCQVAPAVRAIAETG----- 104  
 DB 457 -VSAALIGREYNIALHYPTPRALAEALAACTEVAQNETARNEAPGEPVAVAMAC 514  
 QY 105 -LAPGATTPKO---LAEGLRRL-----GFDEVFTLFGADLTMEGSELLRLTEH 152  
 DB 515 RLPGVSTPEEFWELLSEGRDAVAGLPTDRGMD--LDSLFRPDP--RSGT----- 561  
 QY 153 LEAPHSDPELPMFTSCCPGWIAHLEKYPRLIPYSSCKPQMLAAMVASYAEKKGI 212  
 DB 562 -AHORGGFTLEATFADPAFFGM-----SPREALAVDPQORLMLELSWEVLERAGI 611  
 QY 213 APKDWVAVSIMPCRKQSEADRDWFCVADPPTLRQLDHVTITVELGNIFKRGINLAELP 272  
 DB 612 PPTSL-----QASP-----IGVVGILIPQYGRPLAE-- 638  
 QY 273 EGEDNPMVGSSGAGVLEGTTCGWEALRTAYELFTGTPRLSLSEV----- 321  
 DB 639 -----GGEVEGYLMTGITTSV--ASGRIAVTL--GLEGPALISVDTACSSSLVAVHLA 687  
 QY 322 -----RGMDGIKET-NITWVPAPGSKFEELKHAAARAEAAAGTGPPLAMDGAFPTS 375  
 DB 688 COSLRGESSLAAMGVTVMPTPGLMVDPSMNSLADGRCKAASA-----GANFGM 740  
 QY 376 EDGRGGITTLRVAANGVANGAKLITTKQAGEAKYDFEYI--MACPAGCVGGGGOPRSTDKA 434  
 DB 741 AEGAGMILLE-RLSDARRNGHPVLAIRGTAVNSDGSNGLSAP-----NGRAQYRVYQQA 795  
 QY 435 ITQR--QQAALYNIDESTLRSHEHPSIRELYDTY--LGEF--LGHKAHELLHTHYVA 487  
 DB 796 LAESGLPADIDAVEAHGTGRLGDPTEARALEAYGRDREOPHLGYSKSNLGHQAAA 855  
 QY 488 G 488  
 DB 856 G 856

RESULT 6  
 US-09-036-987A-4

Sequence 4, Application US/09036987A  
 Patent No. 6143526

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.  
 APPLICANT: Broughton, Mary C.  
 APPLICANT: Crawford, Kathryn P.  
 APPLICANT: Madduri, Krishnamurthy  
 APPLICANT: Merlo, Donald J.  
 APPLICANT: Treadway, Patil J.  
 APPLICANT: Turner, Jan R.  
 APPLICANT: Waldron, Clive  
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
 TITLE OF INVENTION: Production  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dow Agrosciences LLC Patent Department  
 STREET: 9330 Zionsville Road  
 CITY: Indianapolis  
 STATE: Indiana

COUNTRY: USA  
 ZIP: 46268  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/036,987A  
 FILING DATE: 09-MAR-1998  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stuart, Donald R.  
 REGISTRATION NUMBER: 28,479  
 REFERENCE/DOCKET NUMBER: 50,608  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (317)337-4816  
 TELEFAX: (317)337-4847  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3170 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-036-987A-4

Query Match 3.9%; Score 100.5; DB 4; Length 3170;  
 Best Local Similarity 21.3%; Pred. No. 2.6;  
 Matches 116; Conservative 64; Mismatches 190; Indels 175; Gaps 33;

43 EAPARRIGNVACAAAPAAAPLASHVQOALAEIAKPKDDPTRKH-----VCQVAPAVR 97  
 DB 1433 EVPAARAKAMANGPAEAGG--SPFARNLAELPEAQ--RRHELVDLYCAQAVAV--- 1481  
 QY 98 AIAETGLACATTPQALAGLRLGFDEVFTLFGADL-----TIME-- 140  
 DB 1482 -----LGHGREVEQPERAFALG---FDSLMAVVDLNRRLTATGRLPTTYFDLP 1530  
 QY 141 EGSELLHRLTEHL-----EAHPHSDPELPMFTSC----- 170  
 DB 1531 NPALAHLELEELVGDVSAAYTAASAPASDEPIATVANSRPGGAHSPEDLMRLVAG 1590  
 QY 171 -----PGWIAHLEKSY-PDLIPYSSCKSPQMLAAM-----VKSYLEAEKKGA 213  
 DB 1591 TEVIGFPPSDRGMDA--EGLYDPD-----ASRPCTTARMAAGLYDAGEFPADLFGIS 1641  
 QY 214 PKDWVAVSIMPCRKQSEADRDWFCVD--ADPTLRQLDHVTITVELGNIFKRG--INL 268  
 DB 1642 PREALAMD-----POQRLVLEIAWEALERAGIDPLSLKSGGVGYIGAGS---RGYATDV 1693  
 QY 269 AELPEGEDNPMVGSSGAGVLEGTTCGWEALRTAYEL-FTGTPLP-----R 315  
 DB 1694 RQFFE-EAEGYILTGTSASVLSG-----RVAVSFGFEPGPAVTVDTCSSSLVALH 1742  
 QY 316 LSLSEVRG--MDGIKETNITWVPAPGSKFEELKHRAA---RAEAHAHTGTPPLAMDGG 370  
 DB 1743 IACOSLRSGCDLALAGVTVMSTP-EMVEVSRQGLAPDRCKSFASAG-IGMGSG 1800  
 QY 371 AGTSEDRGGITTLRVAANGVANGAKLITTKQAGEAKYDFEYI--MACPAGCVGGGGOPR 429  
 DB 1801 A-----GLLERLSDARRNGHRVLAIVRGSAVNODGSNGLAAP---NGPSOOR 1847  
 QY 430 STDKATQRQAALVNLDEKSTLRSHEHPSIRELYDTYLGEPGLGHKAHELLHTHYVAG 489  
 DB 1848 VINQALAA---NAHLSVDVAV--EAHGT-----GTRLGDP--EQALATAY--GQ 1890  
 QY 490 VEEDK 494  
 DB 1891 ARERD 1895

RESULT 7  
 US-09-036-700-4

```

: Sequence 4, Application US/09370700
: Patent No. 6274350
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard H
: APPLICANT: Broughton, Mary C
: APPLICANT: Crawford, Kathryn P
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Treadway, Patti J
: APPLICANT: Turner, Jan R
: APPLICANT: Walton, Clive
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
: FILE REFERENCE: 50489 Div1
: CURRENT APPLICATION NUMBER: US/09/370,700
: CURRENT FILING DATE: 1999-08-09
: EARLIER APPLICATION NUMBER: US 09/36987
: EARLIER FILING DATE: 1998-03-09
: NUMBER OF SEQ ID NOS: 39
: SOFTWARE: Patentl Ver. 2.0
: SEQ ID NO 4
: LENGTH: 3170
: TYPE: PRT
: ORGANISM: Saccharopolyspora spinosa
: US-09-370-700-4

```

```

Query Match          3.9%; Score 100.5; DB 4; Length 3170;
Best Local Similarity 21.3%; Pred. No. 2.6;
Matches 116; Conservative 64; Mismatches 190; Indels 175; Gaps 33;

```

```

QY 43 EAPARRLGNVACAAAPAAEAPLSHVQALAEALAKPKDDPRKH-----VCQVAPAVRV 97
DB 1433 EVPAARKAMPANGPPEPG-----SPFARNLAELEPEAQ-----RRHETLVLCVQAVTV--- 1481
QY 98 AIAETLGLAPATTPKQALAEGLRLGFEVEFTLEFAL-----TIME-- 140
DB 1482 -----LHGSRREVQPERAFRAIG-----FDSLAAVDLRNRLTATGLRLPTTVEYIP 1530
QY 141 EGSELHLRLTEHL-----EAHPHSDPELPMFTSCC----- 170
DB 1531 NPAALAAHLEELVDVSAAVTASAPASDEPAIVAMSCFPGASPEDMLRLVAAG 1590
QY 171 -----PGWIMLEKSY-PDLIPYVSSCKSPOMLAAM-----VKSTLAEEKGIA 213
DB 1591 TEVIGEPFSDRGMDA--EGLYDPP-----ASRPCTTYARWAGFLYDAGEFDADLFGIS 1641
QY 214 PKDWMVSIIMPCTRKQSEADRMFCVD--ADPTLROLDHYTTTELGNIFEREG--INL 268
DB 1642 PREALAMD-----PQORLVLEIAEALERAAGIDPLSKSGVGTITGAGS---RGYATDV 1693
QY 269 AELPGEWDMPMGVSGAGVLEFGTGVMEALRTAYEL-FTGTPLP-----R 315
DB 1694 RQFPE-EAEGYILIGTSAVLSG-----RVAISFGFEPGPAVTVDTACSSSLVALH 1742
QY 316 LSLSEVRG--MDGIKETNITWVPADGSKFEELLKRAAA--RAEAAAGTPGPLAMDG 370
DB 1743 LACQSLRSGECDLALAGVTVWSTP-EMFVEFSRORGLAPDGRCKSFESAAG-IGWSEAG 1800
QY 371 AGFTSEDRGGTTLTVAANGICNAKKLITKQACAEATDYVEI-MACRAGCGGGCQPR 429
DB 1801 A-----GLLLERLSDAHRNGHRLAVALVRSAGVAVODGASNGLAAP---NGPSQOR 1847
QY 430 STDKATQKQALATNLDEKSTLRSHENPSIRELYDTYLGEPLHKHAHELLHTHYVAGG 489
DB 1848 VINQALA---NAALSASVDVAV--EAHGT-----GTRLGDPF--EQALATATY---CQ 1890
QY 490 WEEKD 494
DB 1891 ARERD 1895

```

```

RESULT 8
US-08-426-125-4
: Sequence 4, Application US/08426125
: Patent No. 5853973

```

```

: GENERAL INFORMATION:
: APPLICANT: Kakefuda, Genichi
: APPLICANT: Ott, Karl-Heinz
: APPLICANT: Kwagh, Jae-Gyu
: APPLICANT: Stockton, Gerald W.
: TITLE OF INVENTION: Structure-Based Designed Herbicide
: TITLE OF INVENTION: Resistant Products
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Darby & Darby
: STREET: 805 Third Avenue
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10022-7513
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentl Release #1.0, Version #1.30
: CURRENT APPLICATION DATA: US/08/426,125
: APPLICATION NUMBER: US/08/426,125
: FILING DATE: 20-APR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Robinson, Joseph
: REGISTRATION NUMBER: 33,448
: REFERENCE/DOCKET NUMBER: 0646/0A674
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212)-527-7783
: TELEFAX: (212)-753-6237
: TELEX: 236687
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 638 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Zea mays
: US-08-426-125-4

```

```

Query Match          3.9%; Score 99.5; DB 2; Length 638;
Best Local Similarity 22.3%; Pred. No. 0.23;
Matches 113; Conservative 56; Mismatches 219; Indels 119; Gaps 23;

```

```

QY 39 LATEAPARRLGNVACAAAPAAEAPLSHVQALAEALAKPKDDPRKHVCQVAPAVRYA 98
DB 1 MATAAASTALTGATTAAAPKARRRAHILATRALAA-----PIR---CSAASPMPRA 50
QY 99 IAEITGLAP-GATTPKQ-----LAEGRLRGFEVEFTLEFADLTMEGSELLHR--LT 150
DB 51 PRAT-PLRPMGCTPDRKAGDILVESLERGVADVAPYAGASMET---HQALTRSPVYA 105
QY 151 EULEHPRHSDPELPMFTSQCPCWIMLEKSYDLPYVSSCKSPOMLAAMKSTLAER- 209
DB 106 NLLFRHEDGE-----AFAASGYARSSGRRGVGVCATSGPGATNIVSALADL 151
QY 210 -----KGIAPKDMV-----WVSIIMPCTRKQSEADRMFCVDADPTLROLDHYITT 254
DB 152 LDSVPMVALTGVPRRMICTDAFQETPIVEYTR--STIKHNLVLYDVDDIPRYVQEAFTL 209
QY 255 VELGNIFERGINIALEDPG-----EWDNPMGVSGAGVLEFGTGVMEALRTAYE 306
DB 210 ASSG-----RPGPVLVDIPKDIQOQMAVPMWDRPMSL-----PGYIARLPKPATE 255
QY 307 LFTGTPLRLSLSEVRGMDGIKETNITWV---PAGSKFEELLKHAARAARAAAGTP 362
DB 256 L-----LEQVLRVIGESRRPVLVYGGCAASG--EEL-----RPFVELTGIP 295
QY 363 GPLAMDGAGFTSEDG-----RGGITLRVAV-----ANGIGNAKKLITKQACGA 407

```

Db 296 VTTTLGLGNFSDDELRLMLGMGTYYANYAVDKADLLALGVFDDRVGKIEAFAS 355  
QY 408 KYDFEIMACPGACVGGGQPRSTDKAIQKROALYNDEKSTLRSHENPSIRELYDT 467  
Db 356 RAKIVHVDIDPAE-IGKNQPHVSTICADVKALQGNALLJESTSKSPDFSGWDELQ 414  
QY 468 YLGE-PLGHR-AHELLHTHYVAGVEE 492  
Db 415 QKREFPLGYKTSNEEIOPQYAIQVDE 441

RESULT 9  
US-08-455-355-4  
Sequence 4, Application US/08455355  
Patent No. 5928937  
GENERAL INFORMATION:  
APPLICANT: Kakefuda, Genichi  
APPLICANT: Ott, Karl-Heinz  
APPLICANT: Kwagh, Jae-Gyu  
APPLICANT: Stockton, Gerald W.  
TITLE OF INVENTION: Structure-Based Designed Herbicide  
TITLE OF INVENTION: Resistant Products  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darryl A Darby  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022-7513  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,355  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Robinson, Joseph  
REGISTRATION NUMBER: 33,448  
REFERENCE/DOCKET NUMBER: 0646/1A674-US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)-527-7783  
TELEFAX: (212)-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 638 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: zea mays  
US-08-455-355-4

Query Match 3.9%; Score 99.5; DB 2; Length 638;  
Best Local Similarity 22.3%; Pred. No. 0.23;  
Matches 113; Conservative 56; Mismatches 219; Indels 119; Gaps 23;

QY 39 LATELAPARLGNVACAAAPAAEAPLSHVQALAEIAKPKDPTKRHVQVAPAVRA 98  
Db 1 MATAAASTALTGATTAAPARRRAHLATRALAA-----PIR---CSAASPMPMA 50  
QY 99 IAEITGLAP-GATTPKO---LAELRLGFEVFDTLFGADLTINEGSELHHR---LT 150  
Db 51 PAPT-PLRPWGPFDPRKAGDILVESLERCGYRDVFAIPGASMEI---HQLTRSPVIA 105  
QY 151 EHLEAPHSDEPLPMTSCCPGWIAMLEKSYPDLPVSSCKSPOMLAAVKSYLEAK- 209

Db 106 NHLFRHEGE-----AFASGVARRSGRGVCATSGPATNLVSAADAL 151  
QY 210 -----KGIAPKDMV-----NYSIMPCIRKROSEADRMPCVADPTLRQLDHVIT 254  
Db 152 LDSVPVAITGOVPRRMITDAPQETPIVEYR--SYTHANLVLDVDDIPRYVDEAPFL 209  
QY 255 VELGNIFKRGINTLAELPEG-----EMDPMGVSGAGVLFGTGCVMEALATAYE 306  
Db 210 ASSG-----RPGVLDVIDPIDIOQMAVPWMDPMSL-----PGYIALRPKPATE 255  
QY 307 LFTGTPPLRLSLSEVRGMDGIEKNTIMV---PAPGSFEBELKRAAAREAAAHGTP 362  
Db 256 L-----LEQVRLVGESRRPVLVYGGCAASG---EBL-----RREVELGTIP 295  
QY 363 GPLANDGAGFTSEDC-----RGITLRAV-----ANGIGNAKKLTITKMOGRA 407  
Db 296 VTTTLGLGNFSDDELRLMLGMGTYYANYAVDKADLLALGVFDDRVGKIEAFAS 355  
QY 408 KYDFEIMACPGACVGGGQPRSTDKAIQKROALYNDEKSTLRSHENPSIRELYDT 467  
Db 356 RAKIVHVDIDPAE-IGKNQPHVSTICADVKALQGNALLJESTSKSPDFSGWDELQ 414  
QY 468 YLGE-PLGHR-AHELLHTHYVAGVEE 492  
Db 415 QKREFPLGYKTSNEEIOPQYAIQVDE 441

RESULT 10  
US-08-804-227C-10  
Sequence 10, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3724 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-10

Query Match 3.9%; Score 99.5; DB 2; Length 3724;  
Best Local Similarity 22.8%; Pred. No. 4.4; 255; Indels 111; Gaps 22;  
Matches 124; Conservative 54; Mismatches 111;





```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4550 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-227C-8

```

```

Query Match      3.9%; Score 99.5; DB 2; Length 4550;
Best Local Similarity 18.5%; Pred. No. 6.2;
Matches 114; Conservative 62; Mismatches 192; Indels 247; Gaps 27;

```

```

QY 2 SALVLRPCAAVSRGSS--CRARQA-----PRAPLAATVVALATLEAPARR---- 48
DB 233 AAVVLRPLDALDGDPPYCVVGVAVGNDGGP-----GLTADREGQEA 278
QY 49 LGNVACAA-----APAEAPLSHYQALAEAKRPDP----- 82
DB 279 VLRAACAQAARVDAEYFVELHGTGTPVGDPEAHALGAHSGRADDELLVGSVKTNI 338
QY 83 -----TRKAVCV-----QVAPAVRAIAETLGLAPGA 109
DB 339 GHLEGAAGIAGLVKAAALCLERETLPGLNFATPSPAIPLDQLKLYOTAABE-LPLAPGG 397
QY 110 TTPKOLAEGIRLRGFDEVEFDTLFGADLTMEGSELLHRLTEHLAHPHSDPELPMFTSC 169
DB 398 A---PLAGVSSSGI-----GCTNCHVYLEHLPSRPP----- 427
QY 170 CPWMIMALEKSYDDLIPYSSCKSPQMLLAAMYK-SYLAEEKGIAPKDMYV----- 220
DB 428 ---AVSVAASLPDVPPLLSARSEGALRAQAVRLGEYVERVAGADPRDVAVSLASTRTLF 483
QY 221 ---SIMPCTRKQS-----EADR-----DW----- 236
DB 484 EHRAVYPCGGRGELVVALGGAAGRVSGVSRGAVPGVGLFTGGAGVGMGRGLYA 543
QY 237 ---FCVDADPTLRDLHV-----ITVELGNIFKERGINLAELPEGEMDPMKVGSGAGV 288
DB 544 GGGVFAEVLDEVLNMGVEVDGRSLRDVMPGDVVDAGAGA-----DAGAGAGAGVGS 595
QY 289 LFTTGGV-----MEALRTAYE-----LFTGTPPLPLSLSEYRGMDGIR 328
DB 596 GSGSVGLLGRTEFAQPALFALFEVALFRALEARGVEVSVLGHSGVGAAYVAGVSLTG 655
QY 329 ETNTWTPAPGSKF-----EELKHKRAAAAE---AAAGTGGPLAMDG 369
DB 656 DA-VRLVVARAGLMGLIPVGGGMKSVGASVYGVVEGLGEWVSAVAANG-PRSVVLSG 713
QY 370 GAG-----FTSEDRGRTLRVAVANG-----LGNKKLITKMQAGEARYDFVEIM 415
DB 714 DVGLESEVVASLMDGDVECRRLDYSHGFHVLMEPVLGERGVESLEFRVAPGVVVS 773
QY 416 ACPAGCVGG--GQP 428
DB 774 GVGSGVVGSGELGDP 788

```

RESULT 13

```

US-08-804-198-2
; Sequence 2, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Roostek, Paul R., Jr.
; TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: PAUL R. CANTRELL, 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4550 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-198-2

```

```

Query Match      3.9%; Score 99.5; DB 2; Length 4550;
Best Local Similarity 18.5%; Pred. No. 6.2;
Matches 114; Conservative 62; Mismatches 192; Indels 247; Gaps 27;

```

```

QY 2 SALVLRPCAAVSRGSS--CRARQA-----PRAPLAATVVALATLEAPARR---- 48
DB 233 AAVVLRPLDALDGDPPYCVVGVAVGNDGGP-----GLTADREGQEA 278
QY 49 LGNVACAA-----APAEAPLSHYQALAEAKRPDP----- 82
DB 279 VLRAACAQAARVDAEYFVELHGTGTPVGDPEAHALGAHSGRADDELLVGSVKTNI 338
QY 83 -----TRKAVCV-----QVAPAVRAIAETLGLAPGA 109
DB 339 GHLEGAAGIAGLVKAAALCLERETLPGLNFATPSPAIPLDQLKLYOTAABE-LPLAPGG 397
QY 110 TTPKOLAEGIRLRGFDEVEFDTLFGADLTMEGSELLHRLTEHLAHPHSDPELPMFTSC 169
DB 398 A---PLAGVSSSGI-----GCTNCHVYLEHLPSRPP----- 427
QY 170 CPWMIMALEKSYDDLIPYSSCKSPQMLLAAMYK-SYLAEEKGIAPKDMYV----- 220
DB 428 ---AVSVAASLPDVPPLLSARSEGALRAQAVRLGEYVERVAGADPRDVAVSLASTRTLF 483
QY 221 ---SIMPCTRKQS-----EADR-----DW----- 236
DB 484 EHRAVYPCGGRGELVVALGGAAGRVSGVSRGAVPGVGLFTGGAGVGMGRGLYA 543
QY 237 ---FCVDADPTLRDLHV-----ITVELGNIFKERGINLAELPEGEMDPMKVGSGAGV 288
DB 544 GGGVFAEVLDEVLNMGVEVDGRSLRDVMPGDVVDAGAGA-----DAGAGAGAGVGS 595

```



Search completed: June 3, 2003, 16:09:43  
Job time : 32 secs

---



Db 4600 EEPFLALAGL-RVTDLPAGETPAK-----VDLDTLHEVAGNDGKHATL----- 4641  
QY 266 INIAELPEGEMDNPMGVSGAGVLEGTGGVMEALRTAYELFTGPRRLSLSEVRGD 325  
Db 4642 LGAADLEED------TVRALDRLRT-LEMAAAPDRDLDEVLSPQ 4684  
QY 326 G-----TKETITVPAFGSKFEELLKHRAAARAAAHGTPGLMDGAGFTSEGRG 381  
Db 4685 ESRRLLENDTARPVESSVPLFAEQVAAAPDAVAVGEG--VSMT-----YELDARSD 4739  
QY 382 ITLRVAVANGLGNAKKLITMKG-----AG-----EAKYDEVELMA 416  
Db 4740 ALARSLVAAAGVGVESPVVALERSPEVLSAFLAVAKAGVFPVDSLMPQARVDAY--VAD 4798  
QY 417 CPAG-----CYGGGGQPRSTDKAITOKRQ-----ALVNLDEKSTL 452  
Db 4799 CGARIADVADRPMSGTLVVSAGLGDSDAIVSGLDADRVAVLPAQVPAAVYMYTSGST 4858  
QY 453 RR-----SHEN-----PSIRELYD----- 466  
Db 4859 GRKGVYTHQNLVDLATDTCMGPTPRVLFPHAPFADASSYEIWPILLNGYVAVPRRS 4918  
QY 467 ---TYLGEPLGHAHELLHTHYVAG 488  
Db 4919 IDATVLRDLIG--AHELTHVHTAG 4941

RESULT 2  
US-09-737-149-4  
Sequence 4, Application US/09737149  
Patent No. US20020077466A1  
GENERAL INFORMATION:  
APPLICANT: Spaderna, Steven K  
APPLICANT: Quinn, Kerry E.  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Muralidhara, Padigaru  
APPLICANT: Spytek, Kimberly A.  
TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-620 CIP  
CURRENT APPLICATION NUMBER: US/09/737,149  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 60/170,564  
PRIOR FILING DATE: 1999-12-14  
PRIOR APPLICATION NUMBER: 60/173,165  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: 60/173,362  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: 60/173,544  
PRIOR FILING DATE: 1999-12-29  
PRIOR APPLICATION NUMBER: 60/174,404  
PRIOR FILING DATE: 2000-01-04  
PRIOR APPLICATION NUMBER: 60/174,962  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: 60/223,929  
PRIOR FILING DATE: 2000-08-09  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 965  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-737-149-4

Query Match 4.7%: Score 122; DB 10; Length 965;  
Best Local Similarity 21.9%: Pred. No. 0.1;  
Matches 120; Conservative 46; Mismatches 189; Indels 192; Gaps 22;

QY 7 KCAAVSTRSSGSCARQVAPRAPLAASVRYALATLEAPARRLGNA--CAAAAPAAEAP 64  
Db 25 QPCGVLARLGGSVRLGALLPRAPLAKAARALRAALAPRLPHNLSELYVAAPPRDP 84  
QY 65 LSHVO---QAL-----AELAKPKDDP--TRKHVCQVAVAPVAVVAIAETLGLAP-GATTP 112

Db 85 ASITRGLCOALVPPGVAALLAPPEARPELLQLHFLAAATETPEVLSLREARAPLGAAPN 144  
QY 113 KQALBGLRLGDEVEDTLEFGADLTMEGSELHRLTLEHAHPHSDPRLMFTSCCG 172  
Db 145 FHL-----OLHMASPLET-----LLDVLAVVLQHNAMEDGLACTODPG 185  
QY 173 WIMLEKSY-----PDLIPVSSCKSPOMLAAMVKSYLEKKGIAFKDMVWVSIMPCTK 228  
Db 186 GLVALMTSAGRRPQVLVLSRDTGDAGLRALAPMAAPVSGEAP----- 231  
QY 229 QSEADNDFCVADPPLRLQLDHYITTYELG-NIFKRGINIAELPCEGEMDNPMGVSGAG 287  
Db 232 -----VPAAVLLGCDIARRRRLVLAAPPQPHW----- 258  
QY 288 VLFGTGGVMEALRTAYELFTGPRRLSLSEVR-----GMDGIR 328  
Db 259 -LIGTP--LPPKALPRA-----GLPGLALSGVAPRPLEAIIHDIVQVLAALSGAAV 310  
QY 329 ETNITVPAV-----GSKFEELLKHRAAARAAAHGTPGLMDGAGFTSEGRG 381  
Db 311 QPKRALLPAPVNGDLOPAGPESPGRFLAFLANTSFQGTGPV-WYTGSS--PDEDGQ-- 366  
QY 382 ITLRVAVANGLGNAKKLITMKGAEAKYDFVEIMACPAQ--CYGGGGQPRSTDKAITOKR 439  
Db 367 -----CPAGQLCL-----DQGTNDSAVLDAI 387  
QY 440 QAAVNLDEKSTLRR-----SHENPSIRELY-----DTYLGE 471  
Db 388 FAALANGAPRALRKCCYGCIDLERLADTFFDEFLYVGKXGALRDGHWGLVGD 447  
QY 472 PLGKHAH 478  
Db 448 LLAGRAH 454

RESULT 3  
US-09-737-149-6  
Sequence 6, Application US/09737149  
Patent No. US20020077466A1  
GENERAL INFORMATION:  
APPLICANT: Spaderna, Steven K  
APPLICANT: Quinn, Kerry E.  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Muralidhara, Padigaru  
APPLICANT: Spytek, Kimberly A.  
TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-620 CIP  
CURRENT APPLICATION NUMBER: US/09/737,149  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 60/170,564  
PRIOR FILING DATE: 1999-12-14  
PRIOR APPLICATION NUMBER: 60/173,165  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: 60/173,362  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: 60/173,544  
PRIOR FILING DATE: 1999-12-29  
PRIOR APPLICATION NUMBER: 60/174,404  
PRIOR FILING DATE: 2000-01-04  
PRIOR APPLICATION NUMBER: 60/174,962  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: 60/223,929  
PRIOR FILING DATE: 2000-08-09  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 971  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-737-149-6

Query Match 4.7%: Score 122; DB 10; Length 971;  
Best Local Similarity 21.9%: Pred. No. 0.1;

	Matches	120: Conservative	46: Mismatches	189: Indels	192: Gaps	22:
QY	7	KPCAASVIRGSSCCARQAVAPAPLAASTVVALATLEAPARRIGNVA--CAAAAPAAEAP	64			
Db	25	QPCGVILARRIGSSVRLGALLPRAPLARAARALAAALAPRLPHLSLELVAAAPAPDP	84			
QY	65	LSHVQ---QAL-----AELAKPPDDP--TAKHYCVQAVAPAVRAIAETLGLAP-GATTP	112			
Db	85	ASLTRGLCQALVPPGVAAALLAPPEARRPELQHLAAATEPVPVLSLRRRARAPLGAAPN	144			
QY	113	KQLAEGRLRIGFDEVFDTLFGADLTIMEGSELLHRLTEHLEAPHPSEDEPLPMFTSCCP	172			
Db	145	FHL-----QLHWASPLET-----LADVLAVALQAHAMVEDVGLACTODPG	185			
QY	173	WIAHLEKSY---FDLLPYVSSCKSPQMLAAWKSYLAEEKGIAPKDMVAVSIMCTRK	228			
Db	186	GLVALMTSRAGRPOLVLDLSRRPTGDAGLRAARLAPMAAPVGGEAP-----	231			
QY	229	QSEADRDWFQVDADPTLQLDHYITTVELG-NIFERGINIAELPEGEWDMVMSGSGAG	287			
Db	232	-----VPAVLLGCDIARARVLEAVPPGPH-----	258			
QY	288	VLFETGGVMEALRTAYELTGTPLRPLSLSEVR-----GMDGIRK	328			
Db	259	-LLGTP--LPRKALPTA-----GLPRGLALGEVAPRPLEAIIHDYQLVARALGSAQV	310			
QY	329	ETNITWPAPE-----GSKFEELIKHPAARAEAAAHGTGRLPAMDGCGFTSGDRGG	381			
Db	311	QPKRALLPAPNCCDGLDPAGESPSGRFLARLANTSPFGGRGPV-WVTGSS-PDRDGO--	366			
QY	382	ITLRVAVANGGNKKLITKQAGBAKYDFEIMACPAQ--CVGCGGQPRSDKAITQKR	439			
Db	367	-----CPAGQLCL---DPTINDSATLDAL	387			
QY	440	QALYINDEKSTLR-----SHENPSIRELY-----DTYLGE	471			
Db	388	FAALLANGSAPRALRCQCYCIDILLERLAEDTPFDEELIYLDGDKYKALRDGRMTGLVGD	447			
QY	472	PLGKHAH 478				
Db	448	LLAGRAH 454				
RESULT 4						
US-10-106-534-2						
; Sequence 2, Application US/10106534						
; Patent No. US2002018668A1						
; GENERAL INFORMATION:						
; APPLICANT: Curtis, Rory A.J.						
; TITLE OF INVENTION: 14691, A Human Glutamate Receptor Family						
; FILE REFERENCE: MP101-042P1RM						
; CURRENT APPLICATION NUMBER: US/10/106.534						
; CURRENT FILING DATE: 2002-03-26						
; PRIOR APPLICATION NUMBER: 60/279,086						
; FILING DATE: 2001-03-27						
; NUMBER OF SEQ ID NOS: 7						
; SOFTWARE: FastSeq for Windows Version 4.0						
; SEQ ID NO 2						
; LENGTH: 998						
; TYPE: PRT						
; ORGANISM: homo sapiens						
US-10-106-534-2						
Query Match 4.7%; Score 120; DB 9; Length 998;						
Best Local Similarity 22.3%; Pred. No. 0.16;						
Matches 129; Conservative 45; Mismatches 215; Indels 190; Gaps 24;						
QY	7	KPCAASVIRGSSCCARQAVAPAPLAASVVALATLEAPARRIGNVA--CAAAAPAAEAP	64			
Db	25	QPCGVILARRIGSSVRLGALLPRAPLARAARALAAALAPRLPHLSLELVAAAPAPDP	84			

OY	65	LSHWQALAEIAKPKDDPTKRVHCYQA-PAVRAVAIAETTLGLAPATTPKOLAEGLRLG	123
Db	85	ASLRGICLOALVPP-----GVAALAFPEARPELTLHFLAAATETP--VLSTLRR--	133
OY	124	FDEYFDTLFGADLTINEGSELLNRTLHELEAHPHSDPELPEFTSCCPNMIMLEKSY--	181
Db	134	--EARAFL-GAPLPHASPELTLDVLYAVVLQAHMEDVGLACITQDPGGVALMTSRAG	190
OY	182	--PDLIPVYSSCKSPOMMIAAMVYSYLAIEKKQIARADWVMSIMPCTRKSEADRDWFCV	239
Db	191	RPQVLVIDLSRDRDQDAGLRARLAPMAAPVGEAP-----	225
OY	240	DADPTLRDLHYITTVTELG-NIFKERGINLALPEGEWDPWVGSGAGVLTGTTGYME	298
Db	226	-----VPAAVLLGCDIARARVLEAVPPPHW-----ILGTG--LPP	260
OY	299	AALRAVLELFTGTPLPRLSLSEVR-----GMDGIKETNITVWPA--	338
Db	261	KALPTA-----GLRPGLLALGEVARNPLLEAIHDIQVLVARALGSAAOVQPRALLPAV	315
OY	339	-----GSKFELLRHRAARAEMAANHGP	362
Db	316	NCGLDLPAGPSPGFLARFLAFLANLSPFGQRTGPRVWYTGSSQVNMSHNFKVMSLRPGRGAP	375
OY	363	GPLAWDGAAGFTSEDR-----GGITLLEVAANGLGNKKL--ITKMOAGEAKYDFV--	412
Db	376	---AMATVGSW--RRCQDLDERGGSASARPPRPGQAQWPKLRVYLTLE-----HFFVFAK	425
OY	413	---ETMACPAG--CVGSGGGQPRSTDKAITQKQALVNLDEKSTLNR-----	454
Db	426	DPDEGCGCPAQQLCL---DPGTNDSATLDAFLAALANGSAPRALRKCQGYCIDLLERL	481
OY	455	SHENPSTIRELY-----DYLGEPLGHKHAH	478
Db	482	AEDPPEDEFELVVGDKYKALRGDKWMTGVLGSLLAGRAH	520

```

RESULT 5
US-09-934-070-10
; Sequence 10, Application US/09934070
; Publication No. US20030092004A1
; GENERAL INFORMATION:
; APPLICANT: Lipton, Stuart A.
; APPLICANT: Zhang, Dongxian
; APPLICANT: Chatterton, Jon E.
; APPLICANT: Awobuluyi, Marc
; APPLICANT: Severino, Kevin A.
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
; FILE REFERENCE: P-1J 4900
; CURRENT APPLICATION NUMBER: US/09/934,070
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 897
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-934-070-10

Query Match          4.6%; Score 118; DB 9; Length 897;
Best Local Similarity 21.8%; Pred. No. 0.2;
Matches 128; Conservative 49; Mismatches 207; Indels 202; Gaps 25;

QY      7 KPCAAVSTRGSSCRAROVAPAPLAASVYVALATLEAPARRLGNVA--CAAAAPAAEAP 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      25 QPCGYLARLGSSVRLGAILPRAPPLARARAPALAPALAPRLPHNLSLEYVAAPPAHP 84

QY      65 LSHVO---QAL-----AELAKPKDDP--TERHYCVQYAPAVRYAIAETGLAP-GATTP 112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      85 ASLRGCLQALVPEGVALLAFPEKRPPELLQHLPLAATETRPVLSLRLREARAPLGAVNP 144

QY      113 KQLAEGRLRGFDEVYDTLTFGADLTIMEGSELLRLRLETHLEAHPHSDPLPMTSCCPG 172

```

```
Db 145 FHL-----OLHNASPLET-----LLDLVAVLQAHAMEDVGLACRTQDPG 185
Qy 173 WTAMLEKSY-----PDILPYVSSCKSPQMMLAAMVKSYLEKKGIAPKDMVMYSIMPCTRK 228
Db 186 GLVALMTSRAGRPOLVLDLSRRDTGDAGLRARLAPMAAPVGGAP-----231
Qy 229 QSEADRWFCVADPTLRDLRHVITTVELG-NIFKRGINLAELPEGENDPMWVGSGAG 287
Db 232 -----VPAVLLGCDIARARVLEAVPPCPHM-----258
Qy 288 VLFGTGGVWEALRTAYELFTGTPLRISLSR-----GMDGJK 328
Db 259 -LGTG--LPPKALPTA-----GLPGLLALGEVAPPLAALHDIQVQVLRALGSAQV 310
Qy 329 ENTITWVAP-----GSKFEELKHRAAARAE 355
Db 311 QPKRALLPAPVNCGLDQAPGESPGRPLANTSFOGRTGFWWTGSSQVHMSRFKXWSLR 370
Qy 356 AAAGTTPGLAMDGAGFTSEDR-----GGITLRVAVANGIGNARKL--ITKMOAGEAK 408
Db 371 RDRGAP---AMATVGSW--ROGQDLERPGASARPPPGAGQVMPKRLRVITLLE-----420
Qy 409 YDFV-----ETMACPAG--CVGGGQPPSTDKAITOKROALYNIDKSTLR-----454
Db 421 HPEVFARDPDEDQCPAGOLCTL---DPGTNDSATLIDALFPAALANGSAPRALRKCXYC 476
Qy 455 -----SHENPSIRELY-----DYLGEPLGHKAH 478
Db 477 IDLERLAEDTTPDFELIYVGDKYCALRDRGRTGLVGDILAGRAH 522
```

## RESULT 6

```
US-09-934-070-6
; Sequence 6, Application US/09934070
; Publication No. US2003092004A1
; GENERAL INFORMATION:
; APPLICANT: Lipton, Stuart A.
; APPLICANT: Zhang, Dongxian
; APPLICANT: Chatterton, Jon E.
; APPLICANT: Awobuluyi, Marc
; APPLICANT: Severino, Kevin A.
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
; FILE REFERENCE: P-LJ 4900
; CURRENT APPLICATION NUMBER: US/09/934,070
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1011
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909,
; LOCATION: 910, 931, 932, 933, 934, 935, 936, 937, 938, 943, 944, 945,
; LOCATION: 946, 947, 948, 949
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-934-070-6
```

```
Query Match 4.58; Score 118; DB 9; Length 1011;
Best Local Similarity 21.88; Pred. No. 0.24;
Matches 128; Conservative 49; Mismatches 207; Indels 202; Gaps 25;
```

```
Qy 7 KPCAASVIRSSCAROVAPRAPLAASTVVALATLEAPARLGNVA--CAAAAPAAEAP 64
Db 25 QPCGVLARLGSGVRLGGLPRAPLARARARALARLARLRPHNLSLELVAAFPARDP 84
Qy 65 LSHVQ---QAL-----AELAKPRDP--TRKHVCQVAVAVAYIAETLGLAP-GATTP 112
Db 85 ASLIRGLCQALVPPGVALLAFPEARDELQHLFLAATETPVLSTLRBARAPLGNP 144
Qy 113 KQLAEGIRLGFDEVFTLGLGADLTMEGSELLHRTLEHAFPHSDEPLPMTSSCPG 172
```

```
Db 145 FHL-----OLHNASPLET-----LLDLVAVLQAHAMEDVGLACRTQDPG 185
Qy 173 WTAMLEKSY-----PDILPYVSSCKSPQMMLAAMVKSYLEKKGIAPKDMVMYSIMPCTRK 228
Db 186 GLVALMTSRAGRPOLVLDLSRRDTGDAGLRARLAPMAAPVGGAP-----231
Qy 229 QSEADRWFCVADPTLRDLRHVITTVELG-NIFKRGINLAELPEGENDPMWVGSGAG 287
Db 232 -----VPAVLLGCDIARARVLEAVPPCPHM-----258
Qy 288 VLFGTGGVWEALRTAYELFTGTPLRISLSR-----GMDGJK 328
Db 259 -LGTG--LPPKALPTA-----GLPGLLALGEVAPPLAALHDIQVQVLRALGSAQV 310
Qy 329 ENTITWVAP-----GSKFEELKHRAAARAE 355
Db 311 QPKRALLPAPVNCGLDQAPGESPGRPLANTSFOGRTGFWWTGSSQVHMSRFKXWSLR 370
Qy 356 AAAGTTPGLAMDGAGFTSEDR-----GGITLRVAVANGIGNARKL--ITKMOAGEAK 408
Db 371 RDRGAP---AMATVGSW--ROGQDLERPGASARPPPGAGQVMPKRLRVITLLE-----420
Qy 409 YDFV-----ETMACPAG--CVGGGQPPSTDKAITOKROALYNIDKSTLR-----454
Db 421 HPEVFARDPDEDQCPAGOLCTL---DPGTNDSATLIDALFPAALANGSAPRALRKCXYC 476
Qy 455 -----SHENPSIRELY-----DYLGEPLGHKAH 478
Db 477 IDLERLAEDTTPDFELIYVGDKYCALRDRGRTGLVGDILAGRAH 522
```

## RESULT 7

```
US-09-737-149-33
; Sequence 33, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-737-149-33
```

```
Query Match 4.58; Score 116; DB 10; Length 901;
Best Local Similarity 21.78; Pred. No. 0.31;
Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;
```

```
Qy 7 KPCAASVIRSSCAROVAPRAPLAASTVVALATLEAPARLGNVA--CAAAAPAAEAP 64
```



```

Db      25  QPCGVLARLGGSVRLGALLPRAPRLARARARALAPRALAPRLPHNLSLELVAAAPRARP 84
Oy      65  LSHVQ---QAL-----AELAKPKDP--TRKHVCQVAPAVVAIAFTGLAP-GATTP 112
Db      85  ASLTRGLCOALVPPGVAALLAFEPARPELLQHFLLAAATETVELSLIREARAPLGAAPN 144
Oy      113  KQLAEGURLRGDEVFDTLFGADLTIMEGSSLLRLTTHLEAHNPHSDPRLPFTSCCG 172
Db      145  FHL-----QLHMASPLET-----LLDVLVAVLQAHAWEDGLACRTQDGG 185
Oy      173  WIAMLEKSY----PDLIPYVSSCKSPQMLAAMVKSYLEKKGIAPKDMVMYSIMPCTRK 228
Db      186  GLVALMTSRAGRPQDLVDLSRRDQDAGLRARLAPMAAPVGGAP----- 231
Oy      229  QSEADRFMFCVADPFTLQOLDHVITTVELG-NIFERGINLAEPLGENDNPMKVGSGAG 287
Db      232  -----VPAVLGLGCDIARARVLEAVPPGHM----- 258
Oy      288  VLFCTGGVMEALITAYELFTGTLPRLSLSEVR-----GNDGK 328
Db      259  -LLGTP--LPPKALPTA-----GLPPGLALAGEVARPPLEAAIHDIQVLVARALGSAOV 310
Oy      329  ETNITMVPAP-----GSKFEELKRAA 351
Db      311  QPKRALLPAPVNGSGLQAGRPSPGRFLARFLAINTSFQGRTPGVWYTGSSQYIMSHNFKY 370
Oy      352  ARAEAAHCTPGCLPAMDGAGGFTSEDCR-----GCITLRVAVANGIGNAKKL--ITKQA 404
Db      371  WSLRRDPPGAP---AMATVGSW--RDQQLDLPEPGSAPRPPQGAQVMPKLRVTLLE- 424
Oy      405  GEAKRDFV-----TIMACRPA--QVGGGCGRSPDKAITQKQALYNLDEKSTLR-- 454
Db      425  ---HPEVFARDPDDGCGPAGQCLL---DQGTNSATLDLFAALANGASAPRALRKCC 476
Oy      455  -----SHENPSIRELY-----DTYLEPTLGHKAH 478
Db      477  YGYCIDLLERLAEDTPPDEELTVLDGKYGALRDGRWQVLVDGLLAGRAH 526

RESULT 8
US-09-737-149-8
: Sequence 8, Application US/09737149
: Patent No. US20020077466A1
: GENERAL INFORMATION:
: APPLICANT: Spaderna, Steven K
: APPLICANT: Quinn, Kerry E.
: APPLICANT: Shinkets, Richard A.
: APPLICANT: Muralidhara, Padigaru
: APPLICANT: Spytek, Kimberly A.
: TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
: FILE REFERENCE: 15966-620 CIP
: CURRENT APPLICATION NUMBER: US/09/737,149
: CURRENT FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: 60/170,564
: PRIOR FILING DATE: 1999-12-14
: PRIOR APPLICATION NUMBER: 60/173,165
: PRIOR FILING DATE: 1999-12-27
: PRIOR APPLICATION NUMBER: 60/173,362
: PRIOR FILING DATE: 1999-12-27
: PRIOR APPLICATION NUMBER: 60/173,544
: PRIOR FILING DATE: 1999-12-29
: PRIOR APPLICATION NUMBER: 60/174,404
: PRIOR FILING DATE: 2000-01-04
: PRIOR APPLICATION NUMBER: 60/174,962
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: 60/223,929
: PRIOR FILING DATE: 2000-08-09
: NUMBER OF SEQ ID NOS: 49
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 1043
: TYPE: PRT
: ORGANISM: Homo sapiens

```

[illegible]

Db 4400 PDSASQVAMHR-ROLGAPADELPLADHPRAEATYR-----GHTVEFVPA-----4448  
Qy 65 LSHVQALALAPLAKDDPTKHKVQYAPAVVAIAETLG-LAPGATTPQALAEGLRLG 123  
Db 4449 ---VHQALAEALARNG-----VTFMTVQALANLLSKLAGNDIPIGVAVAGRT--4495  
Qy 124 FDEVDPLFG-----ADLTMEGSELLHRLTE-HLEAHPHSDPELPMFTSCPGW 173  
Db 4496 -DPTLDLIGFVNTLVLRDTLGNPTITDLHRTTRTTLHAFTHODVP-----4543  
Qy 174 IAMEKSYDPLIPYVSSCKSP--OMLAAVKSYLEKKIAPKDMVMSIMPCRKQSE 231  
Db 4544 ---FEKLVEDLAPTRSLARHPPLFQVMTTOSTGRAGAAELPGLEFVLSPGVAAK---4597  
Qy 232 ADQMFQVADPTLRQDHYITVELGNFKERGINALPEGBEMDPMVSGAGVLF 291  
Db 4598 -----VDLD-----LSLSEAVDDDG-----RPAQL---AGLVLA 4623  
Qy 292 TTG---GVNEALRTAYELFTGTPPL-PRLSLSEVRGMDG-----IKETNITMVPAPGS 340  
Db 4624 AADLFEGHTAERLAGYLAIRLAVLPADPGARLDGVDLDGEERIVLTGMNDTAAVPAV 4683  
Qy 341 KPEELKHAARAARAAAHGTPOGPLANDGA--GTSSEDRGGITLRAVAVANGLG 393  
Db 4684 AVPELIERRAAAEPEAGA-----WCGDTHLRYGELNARMLARLALLVERGAG 4731

## RESULT 10

US-10-047-542-75  
; Sequence 75, Application US/10047542  
; Patent No. US20020168367A1  
; GENERAL INFORMATION:  
; APPLICANT: LARRICK, JAMES W.  
; TITLE OF INVENTION: NOVEL IMMUNOADHESINS FOR TREATING AND PREVENTING VIRAL  
; FILE REFERENCE: 030905.0004.CIP1  
; CURRENT APPLICATION NUMBER: US/10/047,542  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/US01/13932  
; PRIOR FILING DATE: 2001-04-28  
; PRIOR APPLICATION NUMBER: 60/200,298  
; PRIOR FILING DATE: 2000-04-28  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 75  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Oryctolagus cuniculus  
US-10-047-542-75

Query Match 4.3%: Score 111; DB 9; Length 912;

Best Local Similarity 20.9%; Pred. No. 0.84; Indels 166; Gaps 26;

Matches 114; Conservative 56; Mismatches 210;

Qy 18 SCRAQYAPRAPLAASTVRYALALE-----APARRIGYVACAAAPAAE-----A 63  
Db 138 SCRAVPGAGPGRSLTLTLRGAQELIRSFAGEPARAGAVLTAVLARRDHGAFSCRA 197  
Qy 64 PLSHVQALAE-----ELAKPKDDPTKHKVQYAPAVVAIAETL-----GL 105  
Db 198 ELDRPQGLALFENSSAPROLMTYALPLDSPR-----LLAPRVLEVDQSLSVCTLDGL 251  
Qy 106 APGATTPKOLAEGLRRLGFEVDFPTLFG-----ADLTMEEG-SELLHRLT---EHLEA 155  
Db 252 FPASACGVHIALGKRLN-PEV--TLEGDAIVATATVAAEEGKOLVCAVTLGGERRES 308  
Qy 156 HPH---SDEPLPMFTSCPGWIMALEKSYDPLIPYVSSCKSPOMLAAVKSYLEKKGI 212  
Db 309 RENTVYVSPARPLT-----LSEPSAPBGKLVYVTCAGARALVTL-----EGV 352  
Qy 213 APKDMVMSIMPCRKQSEADRMVFCVADPTLRQDHYIT---TVELGNFKERGINLA 269

Db 353 PAAPAGPOALOENFNASRSDDGSRFFC---DATLELDGETLKSNGSAELRYLAPR-LDDA 408  
Qy 270 ELPEGEKNDPMG-----VSGAGVLFETGTGVMALARTAYELF 308  
Db 409 DCPR-SWTWPEGEQGLRCEARGNPPPAVHCARSDCGAVLALGLDPVTRALAGTY----463  
Qy 309 TGTPPLRLSLSEYRGMDGIKETNITMVPAP-----GSKPEELKHAARAARAAAGTPTG 363  
Db 464 -----RCTAANVG--EAVVDVYLYEYAPALDSVGCPEVYVLTGSTEASLSCVANGVP 516  
Qy 364 PLAMDGAGFTSBDGRGT-----TLRAVAVANGLSNAKLLITKMQAGEAKYFVEI 414  
Db 517 PSV--SCVRROADVIEGLLVAREHAGTYRCEAINARALAKVAVATVEYGPS-----FEE 570  
Qy 415 MACPAG-----CVGCGG-----OPRSTDKAITOK 438  
Db 571 RSCPSTMTWVGSSEQLFCEVECKPOPSVOCVSGASEGLLPLAPLNPSPSDPSVPRD 630  
Qy 439 ROAALY 444  
Db 631 LARGIV 636

## RESULT 11

US-09-815-242-10258  
; Sequence 10258, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trewick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10258  
; LENGTH: 857  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-815-242-10258

Query Match 4.3%: Score 109.5; DB 10; Length 857;

Best Local Similarity 21.2%; Pred. No. 1; Indels 91; Gaps 19;

Matches 86; Conservative 67; Mismatches 162;

Qy 108 GATTPKOLAGLR-LGDFEVFTLFGADLTMEGSELLHRLTEHLEAHP---SDEPL 163  
Db 312 GATLDBRYIETKDALLERFQKVFVAPDSV-EDTIALIRGLKEVYELHNHQTORAI 370  
Qy 164 PMFTSCPGWIMALEKSYDPLIPYVSSCKSPOM-----MLA 199

Db 371 VAAATLSHRYIA--DRQLPDKAIDLIDEAASSIRMOIDSKPEELRIIDRIITOLKLEQQA 428  
OY 200 AMVKSYLAERKGIAPKDMVMSIMPCTRKSEADRMFCVAD---PTLR-QLDHVIT 254  
Db 429 LMKESDEASKRL--DMINELSDKERYSLEEBEMKAKESLSGTOTIKALEAKIA 485  
OY 255 VE---LGNIFKRGINIAELPEGEWDMVGSAGVLEFCTT---GCVMAALRTAY 305  
Db 486 IQARVGDALMSLEQYKIPLEKQL-----EATQJEGKTMRLRKVTDAIELAEVL 540  
OY 306 ELFTGTPRLSLSEYKMGDKIKETNITVPAPGSEFELLKRAAAREAA----- 356  
Db 541 ARMTGIPVSRMSESE-----REKLIR-----EQLHHRVIGQNEAVDAVSNAIR 585  
OY 357 -AAHGTGGLAMDGAGFTSEDRGIGTLRAVANGLGNAKKLITKMGAGE--AKYDFVE 413  
Db 586 RSRAGLADPNRPIGSLFLGPTGVGKTELCKALANMFSDAMVHIDSEMEKHSYR 645  
OY 414 IMACPAGCVG---GGGQPRSTDKAITOKROALYNLDEKSTLRSH 456  
Db 646 LVGAPPGVYGYEGGVLTEA-----VRRRPYSVILLDE---VEKAI 683

## RESULT 12

US-09-934-070-8  
; Sequence 8, Application US/09934070  
; Publication No. US20030092004A1  
; GENERAL INFORMATION:  
; APPLICANT: Lipton, Stuart A.  
; APPLICANT: Zhang, Dongxian  
; APPLICANT: Chatterton, Jon E.  
; APPLICANT: Awobuluyi, Marc  
; APPLICANT: Sevarino, Kevin A.  
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS  
; FILE REFERENCE: P-12 4900  
; CURRENT APPLICATION NUMBER: US/09/934,070  
; CURRENT FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 1005  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-934-070-8

Query Match 4.2%; Score 108; DB 9; Length 1005;

Best Local Similarity 22.1%; Pred. No. 1.7;

Matches 121; Conservative 52; Mismatches 177; Indels 198; Gaps 28;

OY 7 KCAAIVSIRSSCARQVAPRAPIAATVRAVALTEAPARLGNVACAAAPAAAPIS 66  
Db 27 QCGCVTRAGASVRLAALPRAPARARVLAALAT---PSPLPH----- 68  
OY 67 HVOQALAEIAKPKRDEPT--RKHYCVQAVAVRAVIAETTLGAPGATTPQIAEGLRLGE 124  
Db 69 NLSLELVAVASPTBPASLARGLCQYLA-----PGVV----- 101  
OY 125 DEVEFTLFGADLTIMESSELRLTEHLEAHPSDEPL-----PMFTSCP--- 171  
Db 102 -----ASTTFEAREPL--RLQFLAA--ATETPVLSVLRREVRAPLGRRTPFHL 148  
OY 172 --GMIAMLEKSYPDLLPYVSSCKSPOMMLAAVKSYLEAKKGIAPKDMVMSIMPCTRKQ 229  
Db 149 QIDMASPLETILDIIVLSVLRANAMEDI---ALVLCVRRPSGL-----VLTMTSR 196  
OY 230 SEADRMFCVADPTLRQDHYITVTELGNIKEKGINIAELPEGEWDMVGSAGVYL 289  
Db 197 SQAPK--FYLD-----LSQIDS-----GNDSLRATLALGTEGG-----GTPVSAAYL 238  
OY 290 FCTTGG---VMEALRTAYELFTGTPRLSLSEV---RQMDGIFETNITMVPADGSKF 342  
Db 239 LCCSTAHAEVLEAA--PPSPOMLIGTPLPAEALPKTGLPPGVLVLTGETGCPISLEAAVHDM 297

OY 343 EEL-----LKHRAAAREAA-----AHCTPGPLA-----W 367  
Db 298 VELVRAALSSMLMHPERALLPAAVNCEDLKTGSSSESTARTLARWLSPTSPOGRGAYW 357  
OY 368 DQAGAGTS-----EDGR-----GCTTLRAVAVANGIGNAK 396  
Db 358 VAGSSQVHVSRRFKWLSLRDPLGAPANATVGSMDGQIDFQGAALRVPSPSGTQARP 417  
OY 397 KL--ITKMGAGEAKYDFV-----EIMACPAG--CYGGGGQPRSTDKAITOKROALYNL 446  
Db 418 KLRVYLVE-----HPFVETRESDEGQCPAQQLC---DPGTNDSARLDLFTALENG 468  
OY 447 DEKSTLR 454  
Db 469 SVPRILRR 476

## RESULT 13

US-10-188-246-12  
; Sequence 12, Application US/10188246  
; Publication No. US20030087274A1  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David W.,  
; APPLICANT: Boldog, Ferenc,  
; APPLICANT: Casman, Stacie,  
; APPLICANT: Edinger, Shlomil,  
; APPLICANT: Gerlach, Valerie,  
; APPLICANT: Gorman, Linda,  
; APPLICANT: Li, Li,  
; APPLICANT: Malyankar, Uriel,  
; APPLICANT: Patursajan, Meera,  
; APPLICANT: Peyman, John,  
; APPLICANT: Shenoy, Suresh,  
; APPLICANT: Shimkets, Richard,  
; APPLICANT: Vernet, Corine A. M.,  
; APPLICANT: Voss, Edward  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS  
; FILE REFERENCE: 21402-397B US  
; CURRENT APPLICATION NUMBER: US/10/188,246  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: 60/303046  
; PRIOR FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 60/304502  
; PRIOR FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 60/305011  
; PRIOR FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: 60/306085  
; PRIOR FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: 60/326981  
; PRIOR FILING DATE: 2001-10-14  
; PRIOR APPLICATION NUMBER: 60/360923  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: 60/363636  
; PRIOR FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/373063  
; PRIOR FILING DATE: 2002-04-16  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Custom  
; SEQ ID NO 12  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-246-12

Query Match 4.2%; Score 107.5; DB 9; Length 450;

Best Local Similarity 20.5%; Pred. No. 0.61;

Matches 92; Conservative 53; Mismatches 152; Indels 151; Gaps 20;

OY 17 SSCRAQVAPRAVL-AAATVRAVALA-TLEAP-----ARRLGVACAAAPAAAPAEPL 65  
Db 15 SACR-----PRDLLEAAVYLRAAGAPVPSPGGGGGGGGRTLLQAAGAAVPAAYVR 69  
OY 66 SHVOALAEIAKPKRDPTRKHYCVQV-----APAVRAVAL-----AETL-GLAPGAT 110

```

Db      70 ARAARRAGSGFRNGSVPHRPMMSLYRSLAGRAPAGAAYASGSHGADTITGFTDQAT 129
QY      111 TPQOLAEGLRRLCFD-----EVEFTLFGADLTITMEBSSELHRLTHELEAHPHS--DEPLPM 165
Db      130 ODSAAETGSGFLFEDVSSLNDADDEVGAELRLVLRGSP-----ESGPGSMTSPPLT 180
QY      166 FTSCCGMTAMLEKSPDPLIPYSSCKSPOMMLAAMVKSYLEAKKGIAPDMVMSIMPC 225
Db      181 LLSCTG-----AARAPRLYSRAEPLVGRW-----EAFVADAMRR 219
QY      226 TRKQSEADRMFCVADPTLRQLDHYITVELGNIFKRGINLAELPEGEHMDPMGVSG 285
Db      220 HREPRPPRA-FCL-----LLRAVAGVPSPPLV--- 247
QY      286 AGVLFGTGGVMEALRTAELFTGTPRLSLSEVRGMGDIKETITMYPAGSKFEEL 345
Db      248 RRLGFGWPGG-----GGSAAEERAVLVSSRTQRESL 280
QY      346 LKH-RAAARAEMAA-----HGTGCPPLANDG-----AGFTSEDRGGITLRY 386
Db      281 FREIRAOARALGAALASEPLPDCTGTASPRAYIGRRRRRTALAGRTAQSGG----- 335
QY      387 AVANGIGNAKKLITKMOAGEAKYDFEI 414
Db      336 --GAGRGHGRGRSRSRKPLHYDFEKL 361

```

## RESULT 14

```

US-09-934-070-2
; Sequence 2, Application US/09934070
; Publication No. US20030092004A1
; GENERAL INFORMATION:
; APPLICANT: Lipton, Stuart A.
; APPLICANT: Zhang, Dongxiam
; APPLICANT: Chatterton, Jon E.
; APPLICANT: Awobuluyi, Marc
; APPLICANT: Severino, Kevin A.
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
; FILE REFERENCE: P-LJ 4900
; CURRENT APPLICATION NUMBER: US/09/934,070
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 987
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-934-070-2

```

```

Query Match      4.2%: Score 107.5; DB 9; Length 987;
Best Local Similarity 22.0%: Pred. No. 1.9;
Matches 114; Conservative 48; Mismatches 200; Indels 157; Gaps 23;

```

```

QY      7 KCAAVSIRGSSCRARQVAPRAPLAATVSTVALATLEAPARRLGNVCAAAAPAAEAPLS 66
Db      27 QPCRVPTRAGASVRLAALLPRAPARARVLAALAT---PAPRLPH----- 68
QY      67 HVOQALAEIAKPKDDPT--RKHYCVQVAPRAVAIAETLGLAGATTPKOLAESIRLGF 124
Db      69 NLSLELVAVASPTRDPASLARGLCOVLAAP-----GVVASIAPPEARPE-LRLLOF 118
QY      125 ----DEVFDTLFGADLTITMEBSSELHRLTHELEAHPHSDEPLPMFTSCCGMTAMEKS 180
Db      119 LAATEPVPTRPHQLDMASPLETITLDVLSVRAHAMEDIALVLCVRRPGSLVTLMTN 178
QY      181 Y----PDLPYVSSCKSPOMMLAAMVKSYLEAKKGIAPKDMVMSIMPTKQSEADRDW 236
Db      179 HASQAPKFVLDLSRDSRNDLSRAGLALLGALGGGTPPAAV--LLGSGTARAH----- 232
QY      237 FCVADADPTLRQLDHYITVELGNIFKRGINLAELPEGEW--DNPM-----GVSGCA 286
Db      233 -VLEAAP-----PGPOMLGTPLPAEALPTTGPPGV 263

```

```

QY      287 GVLFGTGGVMEALRTAELF-----TGTPL 313
Db      264 LALGETEONSLAEVNDWELVAQALSSNALVHERALLPAYVNCDDLTGSGSEATGRLL 323
QY      314 PR-LTSEVRGMGDIKETITMYPAGSKFEELKHAARAEMAAHGTGPPAW----- 367
Db      324 ARFLGNTSFGQRTG-----AWWTGSSQVHVSRIFFKWSLRDPLGAP--AMATVGS 373
QY      368 --DGGAGFTSEDRGGITLRYAVANGIGNAKKL--ITKMOAGEAKYDFV-----EIMMG 417
Db      374 WDGQGLDFGP-----GAAALRVPSGTOARPKLRVTVLVE-----HPFVFTRESDEGOC 424
QY      418 PAG--CVGGGCGPSTDKAITQRQAALYLNDEKSTLRR 454
Db      425 PAGQLCL----DGTINDSARLDLFPALVAVGSPRLTR 459

```

## RESULT 15

```

US-09-934-070-4
; Sequence 4, Application US/09934070
; Publication No. US20030092004A1
; GENERAL INFORMATION:
; APPLICANT: Lipton, Stuart A.
; APPLICANT: Zhang, Dongxiam
; APPLICANT: Chatterton, Jon E.
; APPLICANT: Awobuluyi, Marc
; APPLICANT: Severino, Kevin A.
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
; FILE REFERENCE: P-LJ 4900
; CURRENT APPLICATION NUMBER: US/09/934,070
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-934-070-4

```

```

Query Match      4.0%: Score 104; DB 9; Length 1002;
Best Local Similarity 22.1%: Pred. No. 3.8;
Matches 117; Conservative 48; Mismatches 201; Indels 164; Gaps 25;

```

```

QY      7 KCAAVSIRGSSCRARQVAPRAPLAATVSTVALATLEAPARRLGNVCAAAAPAAEAPLS 66
Db      27 QPCRVPTRAGASVRLAALLPRAPARARVLAALAT---PAPRLPH----- 68
QY      67 HVOQALAEIAKPKDDPT--RKHYCVQVAP--AVRAVAIAETLC-----LAPGATPKO 114
Db      69 NLSLELVAVASPTRDPASLARGLCOVLAAPGVASIAFPEARPELRLLLOFLAATETP-- 126
QY      115 LAGLIR-----LGFDEVFTLFGADLTITMEBSSELHRLTHELEAHPHSDEPLPMFTSC 169
Db      127 VASVLRREVTAIG---ADTPHQLDMASPLETITLDVLSVRAHAMEDIALVLCRRV 182
QY      170 CPGMTAMLEKSY----PDLPYVSSCKSPOMMLAAMVKSYLEAKKGIAPKDMVMSIMPC 225
Db      183 DPGSLVLTMTNHAQAPKFVLDLSRDSRNDLSRAGLALLGALGGGTPPAAV--LLGC 240
QY      226 TRKQSEADRMFCVADPTLRQLDHYITVELGNIFKRGINLAELPEGEW--DNPM--- 280
Db      241 STARAH-----VLEAAP-----PGPOMLGTPLPAE 267
QY      281 ----GVSGAGVLFGTGGVMEALRTAELF----- 308
Db      268 ALPTTGPPRYVLALETGEONSLAEVNDWELVAQALSSNALVHERALLPAYVNCDDLK 327
QY      309 -----TGTPLR-LTSEVRGMGDIKETITMYPAGSKFEELKHAARAEMAAHGT 361
Db      328 TGSGEATGRLLARFLGNTSFGQRTG-----AWWTGSSQVHVSRIFFKWSLRDPLGA 380
QY      362 PGPLAW-----DGGAGFTSEDRGGITLRYAVANGIGNAKKL--ITKMOAGEAKYDFV 412

```

Db 381 P---AMATVGSWQDQDFOF---GAALRVPSFGTQARPRLRVTLVE-----HPFV 428  
QY 413 -----EIMACPAG--CYGGGGQPRSTDKAITQKROALYNIDKSTLRR 454  
Db 429 FTRESEDEGQCPAGQCL---DPGTNDSARLDALFAALVNGSVPRTLRR 474

Search completed: June 3, 2003, 16:17:40  
Job time : 26 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:34:37 ; Search time 72 seconds

(without alignments)  
919.799 Million cell updates/sec

Title: CAC80065

Perfect score: 2576  
Sequence: 1 MSALVKPCAAVSIKSSCR.....HELLHTHVAGVEKDEKK 497

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456.5	17.7	476	22	Human oxidoreducta
2	361.5	14.0	456	22	Human NADP hydroge
3	361.5	14.0	456	22	Human protein sequ
4	211.5	8.2	241	22	Drosophila melanog
5	144	5.6	4999	23	Ramoplanin biosynt
6	125.5	4.9	748	20	Granulocytic Ehrli
7	125.5	4.9	748	21	Granulocytic Ehrli
8	125.5	4.9	748	21	Granulocytic Ehrli
9	125.5	4.9	748	21	Granulocytic Ehrli
10	122.5	4.8	573	20	Ehrlichia sp. E82.

11	122	4.7	965	22	AAU02197	Human glutamate re
12	122	4.7	971	22	AAU02198	Human glutamate re
13	116	4.5	901	23	AAO18086	Human N-methyl-D-a
14	116	4.5	1043	22	AAU02199	Human glutamate re
15	116	4.5	4999	23	AAO22159	Ramoplanin biosynt
16	114.5	4.4	11096	22	AAE10129	Streptomyces nous
17	114	4.3	912	16	AAW75206	Rabbit telencephal
18	112	4.3	4472	18	AAW22601	Tyrosine synthase
19	110.5	4.3	451	21	AAV74914	Neisseria gonorrhoe
20	110	4.3	577	17	AAW06554	R. capsulatus acet
21	109.5	4.3	857	22	AAU34665	Staphylococcus aur
22	109.5	4.3	857	22	AAU34665	E. coli cellular p
23	109.5	4.3	4572	19	AAW52845	A. mediterranei ri
24	108.5	4.2	1687	23	ABO08801	Yeast fatty acid s
25	108.5	4.2	1687	23	ABO08802	Yeast fatty acid s
26	108.5	4.2	1687	23	ABO08803	Yeast fatty acid s
27	108.5	4.2	1687	23	ABO08804	Yeast fatty acid s
28	108.5	4.2	1687	23	ABO08805	Yeast fatty acid s
29	108.5	4.2	1687	23	ABO08806	Yeast fatty acid s
30	108.5	4.2	1687	23	ABO08807	Yeast fatty acid s
31	107.5	4.2	447	21	AAV74915	Neisseria meningit
32	107.5	4.2	3398	14	AAW44430	eryA region polype
33	107	4.2	901	21	AAW42494	Human ORF2258
34	105.5	4.1	6797	22	AAW31558	Pimaricin biosynth
35	105	4.1	284	22	ABG17245	Novel human diapo
36	104	4.0	421	21	AAW35747	Zea mays protein f
37	104	4.0	2675	21	AAW07564	Protein encoded by
38	104	4.0	5532	21	AAW23752	S. avermectilis ave
39	104	4.0	5532	22	AAW65267	Streptomyces averm
40	103.5	4.0	1799	23	AAW50359	Mouse laminin-15 b
41	103	4.0	401	21	AAW33383	Zea mays protein f
42	102.5	4.0	451	21	AAV74916	Neisseria meningit
43	102	4.0	722	21	AAV78816	Granulocytic Ehrli
44	102	4.0	739	21	AAV78826	Granulocytic Ehrli
45	102	4.0	756	21	AAV78927	Granulocytic Ehrli

#### ALIGNMENTS

RESULT 1	AAW73690	standard; Protein; 476 AA.
ID	AAW73690	
XX	AAW73690:	
XX	11-SEP-2001 (first entry)	
DE	Human oxidoreductase protein ORF-23.	
XX		
KW	Human oxidoreductase protein; ORF; cell proliferative disorder;	
KW	arteriosclerosis; cirrhosis; psoriasis; cancer; endocrine disorder;	
KW	diabetes mellitus; diabetes insipidus; dwarfism; hirsutism; amenorrhea;	
KW	osteoporosis; metabolic disorder; obesity; phenylketonuria;	
KW	hypercholesterolaemia; reproductive disorder; infertility;	
KW	ovulatory defect; menstrual cycle defect; endometriosis;	
KW	polycystic ovary disease; spermatogenesis disruption; impotence;	
KW	neurological disorder; epilepsy; stroke; Alzheimer's disease;	
KW	Huntington's disease; Parkinson's disease; Creutzfeldt-Jakob disease;	
KW	meningitis; cerebral palsy; muscular dystrophy; mood disorder; anxiety;	
KW	schizophrenic disorder; infection; autoimmune disorder;	
KW	inflammatory disorder; acquired immunodeficiency syndrome; AIDS; asthma;	
KW	allergy; Crohn's disease; atopic dermatitis; gout; multiple sclerosis;	
KW	rheumatoid arthritis; ulcerative colitis; drug screening;	
KW	toxicity screening; transgenic animal; SNP detection; gene therapy.	
OS	Homo sapiens.	
PN	WO200144448-A2.	
PD	21-JUN-2001.	
XX		
PF	07-DEC-2000; 2000WO-US33158.	





QY	167	TSCCGGTATAMEKSY-PDLIPVSSCKSPOMMLAAVKSYLEAKGIAPRDMVMSIMPC	225
Db	169	TSACGWTATYAEIVLGRPTIALICTAKSPQOMGSLVKDYFARQNLSEFKIFHIVARC	228
QY	226	TRKQSEADRMFCVADPTL--RQLDHVITTVELGNIPEKRGINIAELPEGEWDPMGV	282
Db	229	YDKRLKLEALQE-----SLPPLAHGSRGADCVLTGSEINQIWEQDLSVRD-----	272
QY	283	GSGAGVILFG-----TTGGVWEALRLATVELFTGTPRLRLSLSEYKMGDGIK	328
Db	273	-AAVDLLEGDLKEDKVTRRHGASSDGHIAHIFRAKAKELFN-----EDVE	316
QY	329	ENITMVPDAPGSKFEELKHKRAARAAAEAAHGTPOGLAMDGGAGFTSEDRGSIITLRVAV	388
Db	317	E--VTRRLRKNDPFEVTELEK-----NGEVLRFEPA	345
QY	389	ANGLEMANAKLITTKMAGAEAKYDVEIMACPACGVGGGQPRS---TDKAITCKROAALY	444
Db	346	AYGFENIQMILKTLKKKGFPEHFVEVLACAGCCLNGRGAQIPDGHADALLRQMGIYA	405
QY	445	NLDEKSTLRRSHENPSIRELYPTTYICEPLGHAHELHHTHY	485
Db	406	DI---PVRRPRESSAHVOELYQEWLEGINSPRAREVLTHTTY	442
RESULT 3			
ID	AAB92498	AAB92498 standard; Protein; 456 AA.	
AC	AAB92498;		
XX	26-JUN-2001	(first entry)	
DE	Human protein sequence SEQ ID NO:10602.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy.		
OS	Homo sapiens.		
PN	EP1074617-A2.		
PD	07-FEB-2001.		
PF	28-JUL-2000; 2000EP-0116126.		
PR	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
PA	(HELI-) HELIX RES INST.		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
DR	WPI: 2001-318749/34.		
PT	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	full-length cDNAs defined in the specification, and for the detection		
PT	and/or diagnosis of the abnormality of the proteins encoded by the		
PT	full-length cDNAs -		
PS	Claim 8; SEQ ID 10602; 2537pp + CD ROM; English.		
XX	The present invention describes primer sets for synthesizing 5602		
CC	full-length cDNAs defined in the specification. Where a primer set		
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
CC	to the complementary strand of a polynucleotide which comprises one of		
CC	the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		

Query Match	14.0%	Score 361.5	DB 22	Length 456
Query Local Similarity	24.3%	Pred. No. 5e-27		
Matches 112	Conservative	76	Mismatches 174	Indels 99
			Gaps 13	
CC	polynucleotide and an oligonucleotide comprising a sequence complementary to a			
CC	sequence and an oligonucleotide comprising a 3'-end sequence, where the			
CC	oligonucleotide which comprises at least 15 nucleotides and the combination of			
CC	the 5'-end sequence/3'-end sequence is selected from those defined in			
CC	the specification. The primer sets can be used in antisense therapy and			
CC	in gene therapy. The primers are useful for synthesizing polynucleotides,			
CC	particularly full-length cDNAs. The primers are also useful for the			
CC	detection and/or diagnosis of the abnormality of the proteins encoded by			
CC	the full-length cDNAs. The primers allow obtaining of the full-length			
CC	cDNAs easily without any specialized methods. AAH0316 to AAH13628 and			
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to			
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632			
CC	represent oligonucleotides, all of which are used in the exemplification			
CC	of the present invention.			
XX	Sequence	456 AA:		
50				
Query Match	14.0%	Score 361.5	DB 22	Length 456
Query Local Similarity	24.3%	Pred. No. 5e-27		
Matches 112	Conservative	76	Mismatches 174	Indels 99
			Gaps 13	
QY	52	VACAAAPAAEAPLSHVQA-----LAEIANKPKDPTRRKHCQVAPARVAIAETGLA	106	
DB	54	LACDSCTMAEBEGVQLSQGNANQDFEVLNKKCDTSKKKVLVAVSCPSLPFAAKFNLS	113	
QY	107	PGATTPKRLAEGRLRGFDEVFDTLFGADLTIMEGSELNRLTHLEAHPHSDEPLPF	166	
DB	114	V-TDASRLCGFLKSLGIVHYVEDTLIADFSILSESQKFEVRRYRQHS-----EERTLPVL	168	
QY	167	TSCCGGWTAMLEKS--PDLIPVSSCKSPOMMLAMVFSYLAERKGIAPKQVMVNSIMPC	225	
DB	169	TSACGGWRYAERYLGRPTTHALCTAKSPQOVGSLVDYFARQOOLSPKTFHYIADPC	228	
QY	226	TRKQSEADRDWFCVADAPTL--RLDHYITVTEVGEINFEKRGIMLAELPEGEWDPNGV	282	
DB	229	YDKRLAELQ-----SLPRLHSGRQADCVLTSGELAIQMEGDDLVSRD-----	272	
QY	283	GSGAGVLFG-----TTGCVMEALRAYELFTPTPLRSLSEVRGMDGK	328	
DB	273	AAVDTLFGDLKEDKVTNRHDGASSDGHLANIFRNAAKELFN-----EDVE	316	
QY	329	ETNITMVPAPSKPFELLKHPAAARAAEAAGCTPPRLAMDGAGFTSEDRGGITLRVAV	388	
DB	317	E-VYTRLRNKKDFQEVTLK-----NCEVLRFPAA	345	
QY	389	ANGLSNAAKLITKMOAGEAKYDVEIMACPACVGGGQPRS---TDKAITQKRALY	444	
DB	346	AYGFNINQMILAKLKKGFFHFVEVLACAGCCLNGRGRAQTPDGHADALALRQMGITA	405	
QY	445	NIDKSTLRSHPNSIRELYDTYVLEPILGRHAHELHTHY	485	
DB	406	DI----PVRRPSSAHVQELVQEMLEGINSPPARAVLHTTY	442	
RESULT 4				
ID	ABB69383			
AC	ABB69383	standard; Protein; 241 AA.		
XX	ABB69383			
XX	26-MAR-2002	(first entry)		
DE	Drosophila melanogaster	polypeptide SEQ ID NO 34941.		
XX	Drosophila	developmental biology; cell signalling; insecticide;		
XX	pharmaceutical.			
OS	Drosophila melanogaster.			
XX	WO200171042-A2.			
PN	27-SEP-2001.			
PD				



```

Db      4554 RDTTHAFTHQDVP-----PEKLVEDLAPTRSLARHPLFQVMMTLQASAD 4599
OY      209 KGGIAPKQWVMSIMPCTRKQSEADQMCVDADPTLRD---DHYITVELGNITKRG 265
Db      4600 EEPFLAAGL-RVYTDLPAGETPAK-----VDLDLTLHVGAGDGMHATL----- 4641
OY      266 INLAELPEGEMDNPMQVSGAGVLFCTTGVMERALTVAYELFTGTPRLSLSEVRGMD 325
Db      4642 LGAADLFEGE-----TVRALADRLIRT-LEMAAAPDDRLDRILEVLSRG 4684
OY      326 G-----IKETNITWVPAPSGKFEELKHNRAAREAAAHGTPGLANDGAGFTSEDRGG 381
Db      4685 ERSRLVEWMDTARPYVESSVPALFAEQYAAAPDAVAVVGEQ-VSWT-----YRELDARSD 4739
OY      382 ILRLVAVANGIGNAKKLITKMO-----AG-----EAKDPEYEMA 416
Db      4740 ALARSLAAGVGVESPVVALERSPEVLSAFLAVANAGVFPVDSLMPQARVDV-VAD 4798
OY      417 CPAG-----CVGGGGQPRSTDKAITOKRQ-----AALYNLDEKSTL 452
Db      4799 CGARLAVADRPMISGLTVVASAGLGGDSAVVSGDLTADRRAVLPAGPYPGAAYIMYTSGST 4858
OY      453 RR-----SHEN-----PSTRELYD----- 466
Db      4859 GBRKGVVTHQNLVDLATDTQWGPTRPRVLFHAPHAFDASSYEIMVPLNGGVVAVPARRS 4918
OY      467 ---TYLGEPLGHKAHELLHETHYAG 488
Db      4919 IDATVLRDLIG-AHELLTHVHTAG 4941

```

## RESULT 6

AAW89272  
ID AAW89272 standard; Protein; 748 AA.

AAW89272;

04-MAR-1999 (first entry)

Granulocytic Ehrlichia protein clone 52.

Granulocytic ehrlichiosis; Ehrlichia sp.; GE protein; infection; tick;  
diagnosis; vaccine; antigenic protein; antibody; immune response.

Ehrlichia sp.

WO9849313-A2.

05-NOV-1998.

24-APR-1998: 98WO-US08265.

25-APR-1997: 97US-0044933.

(AQUI-) AQUILA BIOPHARMACEUTICALS INC.

Beltz GA, Coughlin RT, Murphy CA, Storey J;

WPI: 1999-009432/01.  
N-PSDB: AAW82787.

New nucleic acid from the human granulocytic ehrlichiosis agent -  
PT and related antigenic proteins, vectors, transformed cells and  
antibodies, useful for diagnosis and in protective vaccines

Claim 16; Fig 7; 154pp; English.

The present sequence represents a granulocytic ehrlichia (GE) protein.  
GE nucleic acids, vectors and host cells are used for the recombinant  
production of GE proteins, and also in research to further characterise  
the proteins. GE protein-encoding nucleic acid molecules are detected  
by hybridisation to GE nucleic acid fragments or by using the fragments

as primers for polymerase chain reaction (PCR) amplification. GE  
proteins, their immunogenic fragments, and GE nucleic acid molecules  
encoding them are used to generate an immune response against GE.  
Specifically as (genetic) vaccines, especially to control ehrlichiosis  
in humans and dogs, but also to raise Ab and to study DNA-protein  
interactions. Ab are used to detect GE proteins by forming an immune  
complex in standard assays, and correspondingly GE proteins can detect  
specific antibodies, especially for diagnosis, assessment and prognosis  
of GE infection, or of contamination of biological samples with GE.

Sequence 748 AA:

Query Match 4.9%; Score 125.5; DB 20; Length 748;

Best Local Similarity 23.2%; Pred. No. 0.0055;

Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

```

OY      248 LDHYITVELGNITKRGINLAELPEGEMDNPMQVSGAGVLFCTTGVMERALTVAYEL 307
Db      98 LEHLITTT-EVISVNEE-----ITPPEGK-----KTLTLFEALTSGRKYGVKALIKNSADV 146
OY      308 FTGTPRLSL-----SEVRGMDGIKETNITWVPAPGS----- 340
Db      147 -NASPEPATITLGIOGRCFQSGKAIKHLKRVYEGAHINPTGSMSPLLAAVQANAEASNL 205
OY      341 ---KEFEELKHNRAAREAAAHGTP-----GPLAMDGAGFTSEDRGGIT 383
Db      206 KEANKTYNFLHHRGADLSSTEHTGTPALHATAAGNHRTRAMLLDGAAPATQDANGRTA 265
OY      384 LRYAVANGIGNAKKLITKMOGAEAKYDVEYIMACPGACVGGGGOPRSTDKAITOKROAL 443
Db      266 LHYAANGDGKLYRMIK-----KCPDSC-----QPLCSDMGDYALHE-AL 305
OY      444 Y--NLDEK---STLRSHEMPSIRELYDTYLGEPGLKHAHELLH 482
Db      306 YSDNVTEKCFIKMKESRKLNSSEFGDILNTPQANBDTLH 349

```

## RESULT 7

AAV78923  
ID AAV78923 standard; Protein; 748 AA.

AAV78923;

23-MAY-2000 (first entry)

Granulocytic ehrlichia NT1 protein sequence.

Granulocytic ehrlichia; granulocytic ehrlichiosis; vaccine; NT1;  
prevent; treatment.

Ehrlichia sp.

WO200006744-A1.

10-FEB-2000.

23-OCT-1998: 98WO-US22512.

28-JUL-1998: 98US-0094381.

(AQUI-) AQUILA BIOPHARMACEUTICALS INC.

Murphy CI, Massung RF;

WPI: 2000-195304/17.  
N-PSDB: AAZ92240.

Novel granulocytic ehrlichia nucleic acid molecules, their polypeptides  
useful as vaccines for treating ehrlichiosis in mammals e.g. humans,  
pigs and dogs

Claim 15; Fig 24; 192pp; English.

XX

PT Novel granulocytic ehrlichia nucleic acid molecules, their polypeptides

associated vaccines for creating enthalloids in mammals e.g. humans, pigs and dogs -

PS Claim 15; Fig 26; 192pp; English.

XX This sequence represents the granulocytic ehrlichia (GE) NY3 protein  
 CC sequence. The invention relates to 13 GE genes W11, W12, W13, W14, W1C,  
 CC NY1, W12, NY3, SWED, BOV, EQ, SLOV1, and SLOV2 isolated from 13 different  
 CC GE clones from a dog, a cow, a horse and ten humans. Granulocytic  
 CC ehrlichia is the causative agent of granulocytic ehrlichiosis, an acute  
 CC potentially fatal tick-borne infection. A vaccine comprising a GE nucleic  
 CC acid molecule or the polypeptide that it encodes, is used for producing  
 CC an immune response in a host to prevent granulocytic ehrlichiosis in an  
 CC animal. The protein sequences can be used to detect anti-GE antibodies in  
 CC an animal.

XX Sequence 748 AA;

Query Match 4.9%; Score 125.5; DB 21; Length 748;

Best Local Similarity 23.2%; Pred. No. 0.0055;

Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

OY 248 LDHVTITVELGNIFERGINLAELPEGEMDNPMVGSGVLEGTGGVMEALPTAYEL 307  
 DB 98 LEHLITT-EVISVNEE-----ITTEPK-----KTLTREALTSGKYGVAALIKNSADV 146  
 OY 308 FTGTPPLRLSL-----SEVGMGDIKETNITWPAAGS----- 340  
 DB 147 -NASPEPAITLIGRCRCGSKAIKHLKRVNAGAHINPTGSMPLAAVQAANEASVL 205  
 OY 341 ----KFEELKRRARAARAAAHGTP-----GPLAMDGAGFTSEDRGIGIT 383  
 DB 206 KEANKRVNFFLLHRGADLSSTHTGTPLALATAGNHRMTAMLLDKGAPATORDARGRTA 265  
 OY 384 LRVAVANGIGNAKKILITMKGAEAKYDFEIMACGACGCGGGRSTKATOKROAL 443  
 DB 266 LHIAANAGDGKIRYRIAK-----KCPDSC-----QPLCSMDGTALHE-AL 305  
 OY 444 Y--NLDER---STLRSHENPSIRELYDTYLGEPLGKHAHELH 482  
 DB 306 YSDNTEKCFKMLKESRKHLNSNSFFDGLNTPQEANGDITLH 349

RESULT 10

AAW82490  
 ID AAW82490 standard; Protein; 573 AA.

AAW82490;

AC AAW82490;

XX 02-MAR-1999 (first entry)

XX Ehrlichia sp. E82.1 protein.

XX Granulocytic ehrlichia; GE; E82; tick-borne infection; fatal; vaccine;

XX Immune response; detection; diagnosis; Ehrlichiosis.

XX Ehrlichia sp.

XX WO9849312-A2.

XX 05-NOV-1998.

XX 24-APR-1998; 98WO-US08264.

XX 25-APR-1997; 97US-0044869.

XX (AQUIT-) AQUITA BIOPHARMACEUTICALS INC.

XX Belz G, Coughlin RT, Murphy C, Storey J;

XX WPI; 1999-034663/03.

XX N-PSDB; AAW65142.

XX New isolated granulocytic ehrlichia nucleic acids - used to develop

PT products for use in vaccines for inhibiting Ehrlichiosis and for use

PT in detection and diagnosis

XX Disclosure; Fig 10; 184pp; English.

XX This sequence encodes the E82.1 protein which is associated with and has  
 CC been isolated from HL60 cells infected with Ehrlichia sp. GE is an acute  
 CC potentially fatal tick borne infection and the proteins described in  
 CC this invention can be used in vaccines to elicit a beneficial immune  
 CC response in an animal to GE. They can be used for inhibiting Ehrlichiosis  
 CC in an animal. The products can also be used for detection and diagnosis.

XX Sequence 573 AA;

Query Match 4.8%; Score 122.5; DB 20; Length 573;

Best Local Similarity 26.1%; Pred. No. 0.0072;

Matches 57; Conservative 14; Mismatches 80; Indels 67; Gaps 9;

OY 286 AGVLFGTTGGVNE---ALRTAYELFTGTPRLSLSEVGMGDIKETNITWPAAGSKF 342  
 DB 3 ACAHINTPTGSMSPPLAAVQAANE-----ASNKEAN-----KI 36  
 OY 343 EELKRRARAARAAAHGTP-----GPLAMDGAGFTSEDRGIGITLRVAVA 389  
 DB 37 VNFLHRGADLSSTHTGTPLALATAGNHRMTAMLLDKGAPATORDARGRTALHIAA 96  
 OY 390 NGLNNAKKILITMKGAEAKYDFEIMACGACGCGGGRSTKATOKROALY--NED 447  
 DB 97 NGDGKLYRIAK-----KCPDSC-----QPLCSMDGTALHE-ALYSDNVT 136  
 OY 448 EK---STLRSHENPSIRELYDTYLGEPLGKHAHELH 482  
 DB 137 EKCFKMLKESRKHLNSNSFFDGLNTPQEANGDITLH 174

RESULT 11

AAU02197  
 ID AAU02197 standard; Protein; 965 AA.

AAU02197;

XX 26-SEP-2001 (first entry)

XX Human glutamate receptor-like protein, MEM2.

XX Glutamate receptor; MEM1; therapeutic; diagnostic; MEM2;

XX human; Alzheimer's disease; Parkinson's disease; cancer; nephrology;

XX female reproductive health; lung disorder; brain disorder; schizophrenia;

XX heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;

XX cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;

XX vision-related disorder; neoplastic pathology; MEM7; MEM8.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 1..965 /note= "All X's have been input by indexer; these

XX represent unreadable residues in the specification"

XX WO200144473-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000WO-US33909.

XX 14-DEC-1999; 99US-0170564.

XX 27-DEC-1999; 99US-0173165.

XX 29-DEC-1999; 99US-0173362.

XX 04-JAN-2000; 2000US-9965564.

XX 09-AUG-2000; 2000US-0223929.

XX 13-DEC-2000; 2000US-9965565.

XX (CURA-) CURAGEN CORP.

PI Spaderna SK, Quinn KE, Shinkets RA, Muralidhara P, Spytek KA;  
XX  
DR WPI; 2001-398154/42.  
DR N-PSDB; AAS06333.  
XX  
PT Novel polypeptide comprising members of protein families (e.g.,  
PT seven-pass transmembrane receptor proteins) according to presence of  
PT domains and sequence relatedness are useful for treating or preventing  
XX e.g., Alzheimer's and Parkinson's -  
XX  
PS Claim 1; Fig 8; 162pp; English.

CC The sequence represents the amino acid sequence of glutamate receptor-  
CC like protein, MEN2, selected from a group (MEM1-MEM8) comprising  
CC members of protein families according to the presence of domains and  
CC sequence relatedness, e.g., seven-pass transmembrane receptor protein  
CC (MEM1), glutamate receptor (MEN2-MEM4), potassium channel protein (MEM5),  
CC phosphatase I protein (MEM6), and retino-binding protein (MEM7-MEM8).  
CC The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are  
CC all useful for treating or preventing a pathology associated with (I)  
CC comprising administering (I), (II), or (III) to a subject (preferably a  
CC human). In addition, (I), (II), and (III) may be used to manufacture a  
CC medicament for treating a syndrome associated with a human disease that  
CC is associated with (I). Furthermore, (I) may be used to identify agents  
CC that bind to it, screen modulators of its activity and determine the  
CC presence or predisposition to a disease associated with altered levels of  
CC (I). Disorders for MEM1 include Alzheimer's or Parkinson's disease,  
CC cancer, nephrology, and female reproductive health. Disorders for MEM4  
CC include those involving the lung and/or brain (e.g., schizophrenia). For  
CC MEM5, disorders include heart (arrhythmic disorders) and other muscular  
CC disorders, clotting deficiencies and cobalamin deficiencies (e.g.,  
CC pernicious anemia). Such disorders for MEM6 include diabetes, whereas  
CC disorders for MEM7 and MEM8 include vision-related disorders, cancer,  
CC and other neoplastic pathologies.

50 Sequence 965 AA;

Query Match	Score	DB	Length
4.78;	122;	22;	965;

Matches 122; Conservative 45; Mismatches 187; Indels 194; Gaps 23;

Matches 122; Conservative 45; Mismatches 187; Indels 194; Gaps 23;

[illegible]

OY	439	ROAALYLDKSKLRR-----	SHENSTRELY-----	DTYIG	470
				:	
Db	387	LPALALAGSPRALRCCTCYCIDLERAAEDTPDFDELIVGDSKYGALPDGHWGLVG			446
OY	471	EPYGHKAH			478
		:	:	:	
Db	447	DLLAGRAH			454

RESULT 12  
AAU02198

AAU02198

ID AAU02198 standard; ProteIn; 971 AA

AC AAU02198

DT 26-SEP-2001 (first entry)

Human glutamate receptor-like protein, MEM3

KW Glutamate receptor; MEM1; therapeutic; diagnostic; MEM2;

KW female reproductive health; lung disorder; brain disorder; schizophrenia;

KW cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;  
KW vitamin-related disorders; pathology; MEV7; MEV8

[illegible]

XX  
DN : WO300144473-A3

XX  
PD 31-TIN-2001

XX  
DE 14-DEC-2000:

XX 14-DEC-1999. 0905-012056A  
PB

PR	27-DEC-1999;	99US-0173165
PR	27-DEC-1999;	99US-0173362

PR 29-DEC-1999; 9905-01/3544  
PB 01-JAN-2000; 2000US-8865564

PR 09-AUG-2000; 2000US-0223929;  
PR 13-DEC-2000; 2000US-9966565

XX  
XX  
(CIRPA-) CIRPACEN COOP

XX	Coastal	CR	Cut	WE	ch
PI					

XX  
DB WPT : 2001-309154 /A3

DR N-PSDB; AAS06334.  
XX

PT	Novel polypeptide con
PT	seven-pass transmembr

PT domains and sequence relatedness are useful for treating or preventing Alzheimer's and Parkinson's -

XX  
PS  
Claim 1: English  
162ms

The sequence represents the amino acid sequence of the protein.

CC like protein, MEM3, selected from a group (MEM1-MEM8) comprising  
CC members of protein families according to the presence of domains and  
CC sequence relatedness, e.g., seven-pass transmembrane receptor protein  
CC (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5),  
CC phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8).  
CC The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are  
CC all useful for treating or preventing a pathology associated with (I)  
CC comprising administering (I), (II), or (III) to a subject (preferably a  
CC human). In addition, (I), (II), and (III) may be used to manufacture a  
CC medicament for treating a syndrome associated with a human disease that  
CC is associated with (I). Furthermore, (I) may be used to identify agents  
CC that bind to it, screen modulators of its activity and determine the  
CC presence or predisposition to a disease associated with altered levels of  
CC (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease,  
CC cancer, nephrology, and female reproductive health. Disorders for MEM4  
CC include those involving the lung and/or brain (e.g., schizophrenia). For

CC MEN5, disorders include heart (arrhythmic disorders) and other muscular  
CC disorders, clotting deficiencies and cobalamin deficiencies (e.g.,  
CC pernicious anaemia). Such disorders for MEN6 include diabetes, whereas  
CC disorders for MEN7 and MEN8 include vision-related disorders, cancer,  
CC and other neoplastic pathologies

**SQ Sequence 971 AA;**

Query Match	Score	DB	Length
4.7%	122	22	971
31.0%	122	22	971
31.0%	122	22	971

Matches 120; Conservative 46; Mismatches 189; Indels 192; Gaps 22;

7 КРЧААВСИ

QY	7	KPCAAYSTIRGGSSCARQVAPRAPIAATSVTALATLEAPARLGVA--CAAAAPAAAP	64
Db	25	QPCGLATGGSVRLGALLPRAPLARARARAALAPLPHLSLEIVYAANRAPDP	84
QY	65	LSHWQ---QAL-----AELAKPKDDP--TKRHCVQVAPRAVRAIAETGLAP-GATTP	112
Db	85	ASTRLGLOALPPGVAAILLAPPEARPEPLQLDHLFAATAETPPVSLTRREARAPGAQNP	144
QY	113	KQLAEGLRRLGFDEVFDTLPGADLTIMEGSESLHRLTEHLEAHPSDEPLPMFTSCGP	172
Db	145	FHL-----QLHMASPLET-----LDDVLAVALQAHMMEVGLACTGTDPG	185
QY	173	WIAHLEKSY----FDLIPYVSSCSPOMLLAAMKSYIAEKKGLAPKDMVAVSIIMPCTRK	228

RESULT 13
AA018086
ID AA018086 standard; Protein; 901 AA

AC	AAO18086;
XX	
DT	05-SEP-2002 (first entry)
XX	
DE	Human N-methyl-D-aspartate receptor

KM Human; N-methyl-D-aspartate receptor; NMDA receptor; asthma;  
KM chromosome 19p11.3; genito-urinary system disorder; urinary incontinence;  
KM benign prostate hyperplasia; nervous system disorder; antiasthmatic;  
KM uropathic; noctropic; neuroprotective; gene therapy.

OS	Homo sapiens.
XX	
PN	W0200240538-A2.
XX	
PD	23-MAY-2002.

XX	16-NOV-2001; 2001WO-EP13264
PF	
XX	
PR	17-NOV-2000; 2000US-249273P.

PI ·Kossida S:

DR WPI; 2002-500204/53.  
DR N-PSDB; AAL47419.

PT New human N-methyl-D-aspartate receptor polypeptide for identifying  
 PR modulating agents useful in treating diseases e.g. asthma -  
 XX  
 PS Claim 25; Fig 2; 108pp; English.

CC The present invention provides the protein and coding sequences of the  
CC human N-methyl-D-aspartate (NMDA) receptor. The sequences are useful in  
CC the identification of modulators of NMDA receptor, and in the treatment  
CC of diseases such as asthma, a genito-urinary system disorder such as  
CC urinary incontinence and benign prostatic hyperplasia or a peripheral or  
CC central nervous system disorder. The NMDA gene is found on chromosome  
CC 19p13.3. The present sequence is the protein of the invention.

**SQ Sequence 901 AA;**

Query Match	4.58;	Score 116;	DB 23;	Length 901;
Post Local Simulation	31.79;	Prod No 0.067;		

Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;

```

0Y      7 KPCAANSTRSSCCARQVAPAPAPLAASVVRALATLTLEAPRLNNVA--CAAAAPAAEAP 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     25 QPCGVLAIRLGGSVRLGALLPPAPLARARARALALAPALPRLNLSIELVYAAAPRAPDP 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      65 LSHVO--QAL-----AELAKPKDDP--TRKHVCYQVAPAVRAVAIETIGIAP-GATTP 112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     85 ASIRGTCQALVDPGVAALLAFPEARPELLOHLELAATETPVSLLREARAPLIGAPNP 144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     113 KQLABGLRLGDEVFOTLPGADLTIMESGELLHRLTEHLEAPHSDERLPMFTSCCPG 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     145 FHL-----QLHMASPLET-----LIDVLVAIVQAHAMEVGLALCRTPDGP 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     173 WIAMLEKSY-----BDLIPIYVSSCKSPQWMLAANYKSYLAEEKGIAPKDMVWVSI MPCRK 228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     186 GLVALWTSRAGRPPOLYLDLSRBDTGADGLARLARLAPMAAPVGGEP----- 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     229 QSEADRMFCVDADPTLRLQDHYITVTVELG-NIFKERGINIAELPEGEQMDPMGVSGAG 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     232 -----VPAAVILGGDIARARVLEAVPPGPHM----- 258
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     288 VLFGTGGVMEALATATAYELFTGPRPLPLSISEVR-----GMDGIC 328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     259 -LLGTP-LPRKALPTA-----GLRPGLLALGEVARPLLEAIIHDIVQLVARALGSAQV 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     329 ETNITMVPAP-----GSKREELIKHRAA 351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     311 QPKRALLPAPANNCGDLPAGEPSQGRFLARFLANTSFQGRGTPVAVWTGSSOVHHSRHKV 370
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     352 ARAEAAAGTGPPLAMDGAGFTSEDR-----GGITLRVAVANGIGNAKRL--ITKMQA 404
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     371 WSLRRDRPGAP---AMATVGSW--RDQGLDLEPGASAKRPPPGQAGVWPLRLRVYTLLE- 424
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     405 GEAKYDEV-----ETMACPAG--CVGGGGQPRSTDKAIITOKROAALYNLDEKSTLRR-- 454
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     425 ---HPFEARDDPEDGCGCPAGQGLCT---DPGTNDSATLIDALFAALANGSAPRALRCC 476
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     455 -----SHENPSIRELY-----DTYIGEPLGKHAH 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     477 YGICDILIERLAIEDPPDFELYLVDGKGYALRDRGWTGIVGLDLAGRAH 526

```

## RESULT 14

AA02199 standard; Protein: 1043 AA.  
 AA02199: 26-SEP-2001 (first entry)  
 Human glutamate receptor-like protein, MEM4.  
 Glutamate receptor; MEM1: therapeutic; diagnostic; MEM2: human; Alzheimer's disease; Parkinson's disease; cancer; nephrology; female reproductive health; lung disorder; brain disorder; schizophrenia; heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3: cobalamin deficiency; pernicious anaemia; diabetes; MEM4: MEM5; MEM6: vision-related disorder; neoplastic pathology; MEM7; MEM8.  
 Homo sapiens.  
 WO200144473-A2.  
 21-JUN-2001.  
 14-DEC-2000; 2000MO-US33909.  
 14-DEC-1999; 99US-0170564.  
 27-DEC-1999; 99US-0173165.  
 29-DEC-1999; 99US-0173362.  
 04-JAN-2000; 2000US-9966564.  
 09-AUG-2000; 2000US-0223929.  
 13-DEC-2000; 2000US-9966565.  
 (CURA-) CURAGEN CORP.  
 Spaderna SK, Quinn KE, Shinkens RA, Muralidhara P, Spytek KA; WPI; 2001-398154/42.  
 N-PSDB; AAS06335.  
 Novel polypeptide comprising members of protein families (e.g., seven-pass transmembrane receptor proteins) according to presence of domains and sequence relatedness are useful for treating or preventing, e.g., Alzheimer's and Parkinson's.  
 Claim 1: Fig 18; 162pp; English.  
 The sequence represents the amino acid sequence of glutamate receptor-like protein, MEM4, selected from a group (MEM1-MEM6) comprising members of protein families according to the presence of domains and sequence relatedness, e.g., seven-pass transmembrane receptor protein (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5), phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8). The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are all useful for treating or preventing a pathology associated with (I) comprising administering (I), (II), or (III) to a subject (preferably a human). In addition, (I), (II), and (III) may be used to manufacture a medicament for treating a syndrome associated with a human disease that is associated with (I). Furthermore, (I) may be used to identify agents that bind to it, screen modulators of its activity and determine the presence or predisposition to a disease associated with altered levels of (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease, cancer, nephrology, and female reproductive health. Disorders for MEM4 include those involving the lung and/or brain (e.g., schizophrenia). For MEM5, disorders include heart (arrhythmic disorders) and other muscular disorders, clotting deficiencies and cobalamin deficiencies (e.g., pernicious anaemia). Such disorders for MEM6 include diabetes, whereas disorders for MEM7 and MEM8 include vision-related disorders, cancer, and other neoplastic pathologies.

Query Match 4.5%; Score 116; DB 22; Length 1043;  
 Best Local Similarity 21.7%; Pred. No. 0.085;

Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;  
 QY 7 KPCAAYSIRSCSRRARQVAPRAPLAATVVALATEAPARRIGNVA--CAAAPAAEAP 64  
 Db 25 QPCGVLAIRLGGSTRGLALPLRAPLARARARARARARARARARARARARARARAR 84  
 QY 65 LSHVQ---QAL-----AELAKPRDDP--TRKHVCQVAVAVAVAIETLGLAP-GATTP 112  
 Db 85 ASLTRGICQALVPPGVAALLAFPEARRELLQLHFLAAETETPVLSILRRARAPLAPNP 144  
 QY 113 KQIAEGIRLGFPEVDITLFGADLTMEBSSELLHRLTELEHNPISDEPLPFTSCCPG 172  
 Db 145 FHL-----QLHMASPLET-----LLDVLVAVLQAHMEDVGLALCRQDPG 185  
 QY 173 WIMLEKSY-----PDILPYVSSCKSPQMLAAVKSYLEAKKGIAPKDMVMSIMPCTR 228  
 Db 186 GLVALMTSRAGRPOLVLDLSRDTGDAGLRALAPMAAPVGEAP----- 231  
 QY 229 QSEADRDWFCVADPTLRQLDHYITVTELG-NIFKRGIMLAEPEGEMDNPGVSGAG 287  
 Db 232 -----VPAAVLGCIDIAARRVLEAVPPGPH----- 258  
 QY 288 VLFETTGVMALRTAYELFTGTPLRLSLSEVR-----GMDGTR 328  
 Db 259 -LIGTP--LPPKALPTA---GLPPGLALGEVAPRLEAHIIVQLVARALGSAQV 310  
 QY 329 ETNITWVPAP-----GSKFEELKRRAA 351  
 Db 311 QPKRALLPAPVNGCDLPACPESPGRFLANFLANTSFGCTGVTGWTTGSSQVMSRHKV 370  
 QY 352 ARAEAAAHCTPGPLADGAGFTSEDR-----GITLRVAVANGLSNANKL--ITKMA 404  
 Db 371 WSLRDRPRGAP---AAATVGSW--RDGQDLLEGGASARPPPGAGQVWPKLVVITLE- 424  
 QY 405 GEAKYDFV-----ETMACPAG--CVGGGCGPSTKRAITQKQAAALYNIDEXSTLR-- 454  
 Db 425 ---HPFVFARDDPEGOCFAGQCL---DPGTNDSATLDALFALANCSAPRALRKCC 476  
 QY 455 -----SHENPSIRELY-----DYLGEPLGKHAH 478  
 Db 477 YGICIDILERLAEDTPEDFELYVGDKYCALDGRRTGLVGDLLAGRAH 526  
 RESULT 15  
 AA02159 standard; Protein: 4999 AA.  
 AA02159: 03-OCT-2002 (first entry)  
 Ramoplanin biosynthetic ORF 14 protein.  
 Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic; biosynthesis gene cluster; bioengineering; peptide synthetase module; adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor; chlorinate; lipopeptide.  
 Actinoplanes sp.  
 WO200231155-A2.  
 18-APR-2002.  
 15-OCT-2001; 2001MO-CA01462.  
 13-OCT-2000; 2000US-239924P.  
 12-APR-2001; 2001US-283296P.  
 24-JUL-2001; 2001US-0910813.  
 (ECOP-) ECOPRIA BIOSCIENCES INC.  
 Farnet CM, Zazopoulos E, Staffa A;



XX WPI: 2002-435445/46.  
DR N-PSDB: AAL40781.

XX Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
PT chemically modifying biological molecule that is a substrate for a  
PT polypeptide encoded by a ramoplanin biosynthesis gene cluster -

PS Claim 14: Page 169-186; 212pp; English.

XX  
CC The invention relates to an isolated ramoplanin biosynthetic pathway  
CC polypeptide selected from a polypeptide of open reading frames (ORF)  
CC 1-33. The isolated polypeptides are useful for chemically modifying a  
CC biological molecule that is a substrate for a polypeptide encoded by a  
CC ramoplanin biosynthesis gene cluster, by contacting the biological  
CC molecule with the isolated polypeptide, where the polypeptide chemically  
CC modifies the biological molecule. The method comprises contacting the  
CC biological molecule with at least two different polypeptides encoded by  
CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
CC gene cluster comprising the ORFs is useful as a substrate for  
CC bioengineering of antibiotic structures. An isolated polypeptide or its  
CC encoding nucleic acid sequence is useful for generating derivatives of  
CC ramoplanin, for improving production or for producing variants of other  
CC antibiotics of the peptide class. The isolated polypeptides are useful  
CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain  
CC in conjunction with other peptide synthetase modules and allowing the  
CC incorporation of Thr into a peptide antibiotic precursor, for modifying  
CC fatty acid structure and/or enhancing fatty acid incorporation into the  
CC peptide antibiotic structure, for production of an hydroxyphenylglycine  
CC (HPG)-containing peptide antibiotic, for enhancing secretion of  
CC ramoplanin or its variants and derivatives, for enhancing uptake of  
CC precursors for ramoplanin biosynthesis, for enhancing production of  
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of  
CC a peptide antibiotic precursor, and for designing specific nucleotide  
CC probes and primers for identifying and isolating putative lipopeptide  
CC -producing microorganisms. This sequence represents one of the ORF  
CC proteins of the ramoplanin producing Actinoplanes sp. microorganism of  
CC the invention.

XX  
SQ Sequence 4999 AA:

Query Match 4.5%; Score 116; DB 23; Length 4999;

Best Local Similarity 24.1%; Pred. No. 1.1;

Matches 100; Conservative 32; Mismatches 171; Indels 112; Gaps 20;

OY 8 PCAAVSIRGSSCRAROVA---PRAPLASTVRVALATLEAPARRLGNAVACAAAPRAEAP 64  
DB 4400 PSDAISQVYAHNR-RQLAGAPDELPLADHPRAEATYR-----GHTVEFTVPPA---- 4448  
OY 65 LSHVQOALAEIAKPKDDPTKRVHCYQVAAVVAIETIG-LAPGATTPKQIAEGLRIG 123  
DB 4449 ---VHHQIAELARRNG-----VTVEFTVQALAVLLSKLAGTDIRIGAVAGRT-- 4495  
OY 124 FDEVEFDTLFG-----ADLTMEBSSELHRLTE-HLEAHPHSEDEPLPMFTSCPGW 173  
DB 4496 -DPTIDNLGFEVNTLVLTDLFGNPTITDLHRTDTLHAFTHODV----- 4543  
OY 174 IAMEKSYPDLLIPYSSCKSP--QUMLAAMVSYLAEEKGIAPKDMVWVSIIMCTRKQSE 231  
DB 4544 ---FEKLVEDLAPTRSLARHPLFQVWMTLQSTGRAGEAELPGLIETAVLSPGGVAAK--- 4597  
OY 232 ARDMFCVADPTLRQLDHNVTITVELGNIFKERGINLAELPEGEMDNPMGVSGAGVLF 291  
DB 4598 -----VDLD-----LSLSEAYDDDG-----RPAGL---ACTLVA 4623  
OY 292 TTG-----GYMEALRTAYELFTGTPL-PRLSLSEVGMKG-----IKFTNITWPA 340  
DB 4624 AADLPEHGTAERLAIAGTLARLAVLPADPARLDGVDLDGEERRLVLTGNDTTAAVP 4683  
OY 341 KFEELKRAARAARAANAHTGTPGLAMDGA--GFTSEDRGGITLRVAVANGIG 393  
DB 4684 AVPELIERAARAAPERA---VWCSDTHLRVGEIENARLRRLALIVERGAG 4731

Search completed: June 3, 2003, 16:07:35  
JOB Time : 75 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 16:09:12 ; Search time 42 Seconds

(without alignments)  
1137.591 Million cell updates/sec

Title: CAC80065

Perfect score: 2576

Sequence: 1 MSALVKPCAASVIRSSCR.....HELTHYVAGVEKDEKK 497

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1010.5	39.2	608	C72405	hydrogenase (EC 1.1.1.1)
2	877	34.0	585	D57150	hydrogenase (EC 1.1.1.1)
3	874	32.0	1206	T18557	probable hydrogenase (EC 1.1.1.1)
4	794.5	30.8	606	S13526	hydrogenase (EC 1.1.1.1)
5	743.5	28.9	574	HQC1LP	hydrogenase (EC 1.1.1.1)
6	741	28.8	582	JC6002	hydrogenase (EC 1.1.1.1)
7	741	28.8	582	D96903	hydrogenase (EC 1.1.1.1)
8	633	24.6	421	HQDYL	hydrogenase (EC 1.1.1.1)
9	599.5	23.3	421	HQDYL	hydrogenase (EC 1.1.1.1)
10	512.5	19.9	645	G72256	hydrogenase (EC 1.1.1.1)
11	311.5	12.1	538	T40992	hydrogenase (EC 1.1.1.1)
12	255	9.9	491	S63206	hydrogenase (EC 1.1.1.1)
13	245.5	9.5	450	B97297	hydrogenase (EC 1.1.1.1)
14	232	9.0	469	S31336	hydrogenase (EC 1.1.1.1)
15	138.5	5.4	123	HQDVS	hydrogenase (EC 1.1.1.1)
16	136.5	5.3	301	B72256	hydrogenase (EC 1.1.1.1)
17	130	5.0	124	HQDVS	hydrogenase (EC 1.1.1.1)
18	125.5	4.9	748	T08612	hydrogenase (EC 1.1.1.1)
19	117	4.5	24	S33302	hydrogenase (EC 1.1.1.1)
20	114.5	4.4	884	PN0667	hydrogenase (EC 1.1.1.1)
21	114.5	4.4	2588	T14342	hydrogenase (EC 1.1.1.1)
22	112.5	4.4	913	B97369	hydrogenase (EC 1.1.1.1)
23	112.5	4.4	913	AB2587	hydrogenase (EC 1.1.1.1)
24	112.5	4.3	755	S23441	hydrogenase (EC 1.1.1.1)
25	111.5	4.3	891	JN0867	hydrogenase (EC 1.1.1.1)
26	110.5	4.3	451	C81059	hydrogenase (EC 1.1.1.1)
27	110	4.3	503	F95988	hydrogenase (EC 1.1.1.1)
28	109.5	4.3	857	D35905	hydrogenase (EC 1.1.1.1)
29	109.5	4.3	857	G91060	hydrogenase (EC 1.1.1.1)

30	109.5	4.3	861	2	E85905	heat shock protein
31	109.5	4.3	4735	2	T17463	ribosecyclic polyketide synthase
32	109	4.2	6420	2	T30283	polyketide synthase
33	108.5	4.2	594	1	G81847	diacylglycerol kinase
34	108.5	4.2	1867	2	S61703	probable acid synthase
35	108	4.2	5149	2	F83345	probable non-ribosomal peptide synthetase
36	107.5	4.2	391	2	T35574	probable alanine aminotransferase
37	107	4.2	378	2	B75547	iron-sulfur cofactor
38	107	4.2	458	2	G64507	hypothetical protein
39	106	4.1	1772	2	T36105	probable large glyoxylate oxidase
40	106	4.1	1885	1	JC4086	probable large glyoxylate oxidase
41	105.5	4.1	3491	2	T43231	probable 6-deoxyerythronolide synthase
42	105	4.1	896	2	G81709	translation initiation factor
43	105	4.1	1613	2	T06678	hypothetical protein
44	105	4.1	1621	2	A82255	hypothetical protein
45	104.5	4.1	733	2	T35429	probable fatty acid synthase

#### ALIGNMENTS

##### RESULT 1

C72405

hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Thermotoga maritima (strain C:Species: Thermotoga maritima)

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: C72405

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, G.A.; Richardson, C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; PMID:99287316; PMID:10360571

A:Accession: C72405

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-608 <ARN>

A:Cross-references: GB:AE001705; GB:AE000512; NID:94980694; PIDN:AMD35293.1; PID:9498

A:Experimental source: strain MS88

C:Genetics:

A:Gene: TM0201

C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2(4Fe-4S) homology

C:Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase

F:62,74,77,91/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

F:123,127,130,136/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type

F:174,177,180,227/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:184,217,220,223/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:329,384,528,532/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:532/Binding site: diron cofactor (Cys) #status predicted

Query Match 39.2% Score 1010.5; DB 2; Length 608;

Best Local Similarity 47.9% Pred. No. 1,9e-63;

Matches 216; Conservative 56; Mismatches 112; Indels 67; Gaps 8;

Oy	49	IGNVACAAMP-AAEAPLSIVQOALAEAKPKDPTKHKYGVAVRAIAETGLAP	107
Db	218	IGCGCAAFPTGAIYVENSAYKYLELEKE-----KILVPTASVRAIAIEERYAP	272
Oy	108	GATTPROLAEGRLRGEFEDTLFGADLTIMEGSELRLTLEHLEAHPHS-D-EPLPMF	166
Db	273	GTISTGOMVALRLRGDYVDFINFGADLTIMEGSEFLERLEK-----GDLLEPLMF	325
Oy	167	TSCCPGRIAMLEKSYDILTYVSSCKSPOMMLAAKSYLAEEKGIAPKMVMVSTPCT	226
Db	326	TSCCPGVMNLVERVPELRLTSSASPOQMLSAMKTYFAELGVAPEDIFVSTIMPCT	385
Oy	227	RKQSEADRDVFCVADADPTLRQLDHTVITVELGINIFKRGINLAELPEGEEDNMGVSGA	286
Db	386	AKDEALRKQAMNGVPA---VDVLTTRGLGLIRKRIKPPANLPEEEDALGLSTGA	442
Oy	287	GVLEGTGGVMEALRTAYELFTGTPPLRLSLSEVSGMDIKETNTITMVPAPSKREEL	346
Db	443	ALFVGTGVMEALRTAYELTKGKALPKIVFEVVRGLKGVRAEIDL-----	490

QY 347 KHRRAARAARAAHGTGTPGLAMDAGFTSEDRGRTILRYAVANGGLNAKLITKMOAGE 406  
 Db 491 -----DGGK---KIRIAVHGTAANANLVEKILRRE 517  
 QY 407 AKQDVEIMACPGACVGGGQPRSTDKATQKROALYNDEKSTLRSHENPSIRELYD 466  
 Db 518 VKHAFVEVMACPGCGICGGQPYSPBELIRKRAEAIYITDERMTRKSHENPAIKKRG 577  
 QY 467 TYGEPGLGKHAHELHTHYAGVGEDEKK 497  
 Db 578 EYLEHPLSHKAEHLHTY-----EDSRKK 603

## RESULT 2

hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Desulfovibrio fructosovorans  
 C:Species: Desulfovibrio fructosovorans  
 C>Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 19-May-2000  
 C/Accession: D57150  
 R:Malik, S.; Salmame, I.; De Luca, G.; Rousset, M.; Dermoun, Z.; Belatch, J.P.  
 J. Bacteriol. 177, 2628-2636, 1995  
 A:Title: Characterization of an operon encoding an NADP-reducing hydrogenase in Desulfovibrio  
 A:Reference number: A57150; MUID:95270577; PMID:7751270  
 A/Accession: D57150  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-585 <MAL>  
 A:Cross-references: GB:U07229; NID:9466362; P1DN:AA87057.1; PID:9466366  
 C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2(4Fe-4S) homology  
 C:Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase  
 F:36/52,55,69/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
 F:101,105,108,111/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type N1)  
 F:153,156,159,206/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
 F:163,196,199,202/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
 F:312,367,510,514/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
 F:514/Binding site: diron cofactor (Cys) #status predicted

Query Match 34.0%; Score 877; DB 2; Length 585;  
 Best local similarity 41.4%; Pred. No. 4,6e-54;

Matches 192; Conservative 66; Mismatches 138; Indels 68; Gaps 9;

QY 33 STYRVATLAEADPARRLGNYA-----CAAAP-AAEAPLSHQOALAEIAPKRPDP 82  
 Db 173 SGVNRGFTAVVADAFEM-NIADVCTNCGCVAVCPGALVEHEIYIEVEALANDP--- 228  
 QY 83 TRKHVCQVAVAVVAIAETLGLAPGATTPKOLAEGRLGDEVEDTTLTGADLTMEEG 142  
 Db 229 --RVYVQTAPAAALAGEDLVAPGTSYTGKMAAALRLRGFDHVEPTDPAADLTMEEG 286  
 QY 143 SELLHRLTELEAHPSDEPLPMFTSCCPGIMALEKSYDPLIPYSSCKSPOMLAAVY 202  
 Db 287 SEFLDRRGKHLAG--DTNVKLPILITSCCPGIMVAFEHQFPDMIDVPTASTAPOOMGALA 344  
 QY 203 KSLIAEKKGIAPRDMYVSIIMPCTRKQSEADRMFCVADPTLRQLDHVTITVELGNIF 262  
 Db 345 KTYVADLGIPIREKLIVVSVMPCLAKKYECARPEFSVGNP---DVDIVITTRRELAKLVK 401  
 QY 263 ERGINLAELPEGEMDNPMGVSAGVLFPTGVMALRLTAVELFPTGLPLRLSEVR 322  
 Db 402 RKNIDRGRLPDEPDPAVLGASTGAAPLFGVTGVTAEALRTAELAGETLTKLRFEDVR 461  
 QY 323 GMDIKETNTITWVPAGSKFEELIKHRAARAARAAHGTGTPGLAMDAGFTSEDDGGI 382  
 Db 462 GMDGVKRAKVKY-----GDN 476  
 QY 383 TLRYAANGAGNAKLLTKMOAGEAAYDEVIMACPGACVGGGQPRSTDKATQKROA 441  
 Db 477 ELVIGVAGHAGNARELILKPCGAGET--FHAIEVMAACPGCGICGGQPYNHGDVIELKKRTG 535  
 QY 442 ALVNNDEKSTLRSHENPSIRELYDYLGEPLGKHAHELHTHY 485  
 Db 536 VLYAEDGKPLRKSHENPSIRELYEKLPLSRSHQLLHTHY 579

## RESULT 3

probable hydrogenase (EC 1.18.99.1) - Myctotherus ovalis (fragment)

C:Species: Myctotherus ovalis

C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C/Accession: T18557

R:Ahmanova, A.; Vonken, F.; van Alen, T.; Van Hoek, A.; Boxma, B.; Vogels, G.; Veenh

Nature 396, 527-528, 1998

A:Title: A hydrogenosome with a genome.

A:Reference number: Z18971; MUID:99075329; PMID:9859986

A/Accession: T18557

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1206 <AKH>

A:Cross-references: EMBL:Y16775; NID:e1360694; PID:e1360695; P1DN:CA176373.1

A:Experimental source: specific host Periplaneta americana

C:Function: catalyzes reactions involving the production or consumption of molecule

C:Keywords: hydrogen metabolism; oxidoreductase

Query Match 32.0%; Score 824; DB 2; Length 1206;  
 Best local similarity 42.1%; Pred. No. 6,5e-50;

Matches 179; Conservative 72; Mismatches 120; Indels 54; Gaps 10;

QY 84 RKHVCQVAVAVVAIAETLGLAPGATTPK-OLAEGRLGDE-VEPTLFGADLTMEEG 141  
 Db 241 QORLVCOMASINVAEEFGIRPEKILNETATRLKGSNVFVLDNFSADLTITEE 300  
 QY 142 GSELHRLTEH-----LEAHPSDEPLPMFTSCCPGIMALEKSYDPLIPYSSCKSPOM 196  
 Db 301 GHLEILERLVANVYGKLLGGDHPIDLPMLTSCCPGIMPEIKENYDPLLNLSSTCKSPG 360  
 QY 197 MLAAMKSYIAEK-KGIAPRDYVSIIMPCTRKQSEADRMFCVADPTLRQLDHVTITY 255  
 Db 361 MGALIKGYWAKNIKMPDIYVSIMPTAKAKEREPQ--LRDQEGKYDVTYILTR 418  
 QY 256 ELGINFERGINIAELPEGEMDNPMGVSAGVLFPTGVMALRLTAVELFTG--TPL 313  
 Db 419 ELAKMLKOSINDIAKEPIPFDPVYMESEGTAAVIFGYTGVMALRLTAEVITGRVPR 478  
 QY 314 PRLSEVRGMDIKETNTITWVPAGSKFEEL-KHRAARAARAAHGTGTPGLAMDAGAG 372  
 Db 479 KNLINIEAVRGMGIREA-----GKLENVLDYKKAPE----- 510  
 QY 373 FTSEDRGRTILRYAVANGGLNAKLL-----TKMOAGEAKYDVEIMACPGACVGGGQ 427  
 Db 511 -----GTVAVVLAHGNNAKRYMDIIRKQAKESGKPAHVEVMACPGCGICGGGQ 562  
 QY 428 PRSTDKATQKROALYNDEKSTLRSHENPSIRELYDYLGEPLGKHAHELHTHYVA 487  
 Db 563 PKPTNLEIRAKRQTLTKEMDPLRKSHNPETIKATYENLYLEPRLSHHNLHTYSS 622  
 QY 488 GGVEE 492  
 Db 623 OKVRD 627

## RESULT 4

hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Desulfovibrio vulgaris

N:Alternate names: hydrogenase gamma

C:Species: Desulfovibrio vulgaris

C>Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 20-Apr-2001

C/Accession: S13526

R:Stokermans, J.; van Dongen, W.; Kaan, A.; van den Berg, W.; Veeger, C.

FEMS Microbiol. Lett. 49, 217-222, 1988

A:Title: hyd-gamma, a gene from Desulfovibrio vulgaris (Hildenborough) encodes a poly

A:Reference number: S13526; MUID:89306533; PMID:2663654

A/Accession: S13526

A:Molecule type: DNA

A:Residues: 1-606 <STO>  
A:CROSS-references: EMBL:X57838; NID:g40827; PIDD:CAA0970.1; PID:g40828  
C:Genetics:  
C:Gene: hycC  
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin [2(Fe-4S)] homology  
C:Keywords: 2pe-2S; 4fe-4S; iron; metalloprotein; oxidoreductase  
E:33,45,48,66/Binding site: 2pe-2S cluster (Cys) (covalent) #status predicted  
F:98,102,105,111/Binding site: 4fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type N1)  
F:152,155,158,205/Binding site: 4fe-4S cluster (Cys) (covalent) #status predicted  
F:162,195,198,201/Binding site: 4fe-4S cluster (Cys) (covalent) #status predicted  
F:305,360,507,511/Binding site: 4fe-4S cluster (Cys) (covalent) #status predicted  
F:511/Binding site: diron cofactor (Cys) #status predicted

Query Match 30.8%; Score 794.5; DB 2; Length 606;  
Best Local Similarity 42.4%; Pred. No. 3,1e-48;  
Matches 169; Conservative 61; Mismatches 112; Indels 57; Gaps 6;

DQ 90 QVAVNVAIAETGLIAGLPPTKRLAEGRLEGFDFVTLLPGADLTIMEGSELLHRL 149  
| | | | | :  
DQ 233 QFAAVAVGVGEELGPFGSSVEGOVFTALRLGADVLDITNPADLVIMEGETLIQLR 292  
| | | | | :  
DY 150 TEHLAAHPHSDEPLPMFTSCCPGMIALEKSYDDLPIYVSSCKSPOMLAAMYSLAER 209  
| | | | | :  
DB 293 -----RCGAKPLEPFTSCCPGMVNFKAERHLPILPHVSPTSPOCCLAGAKTYLAT 344  
| | | | | :  
DY 210 KGAPKAMWVSTMPCTRKOSEADRDMFCVDADPTLRQDLHVITTVELGINFERGINLA 269  
| :  
DB 345 MNVAPERMRVVSTLMPCRAKKKEEAARPEFRRG-----VRDVAVVLTRREFALLRREGIDL 401  
| :  
DY 270 ELPEGENDNP-MGVSGSAGVLFEGTGGVMEALRTAVELTPGTPLPRLSIEVRGMGIK 328  
| :  
DB 402 GLPSPPCDPLMGKATANAAYIFGTTGGVMEALRTAVHVLNGELADVELHALRGYENVR 461  
| :  
DY 329 ETNTTWVPADGSKEHEELKHKRAARAARAAAGTFPGPLAMDGAGFTSEDGCGITTLVAV 388  
| :  
DB 462 EAIV-----PL-----GEGNSVAVAV 478  
| :  
DY 389 ANGLGNNAKLTKMAGEBAKYDEVEMACPGCVGGSGOPRS--TDKAITQKRQALNYNL 446  
| :  
DB 479 VHGKAROMEVAVLAKRADHVEVMACPGCGKDGGOGPRSKRAYNPNAARRAFLSL 538  
| :  
DY 447 DEKTLRRSHNPISIRIELYDTYLGEPLGHKAHELHHRY 485  
| :  
DB 539 DAENALROSHNNPLIGRVESFLGPCSNLHRLHHRY 577  
| :

RESULT 5  
HQCLIP  
hydrogenase (EC 1.18.99.1) (Fe) I, periplasmic [validated] - Clostridium pasteurianum  
N:Alternate names: [Fe] hydrogenase  
C:Species: Clostridium pasteurianum  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 15-Sep-2000  
C:Accession: A40330  
R:Meyer, J.; Gagnon, J.  
Biochemistry 30, 9697-9704, 1991  
A>Title: Primary structure of hydrogenase I from Clostridium pasteurianum.  
A:Reference number: A40330; MUID:92002005; PMID:1911757  
A:Accession: A40330  
A:Molecule type: DNA  
A:Residues: 1-574 <ME>  
A:CROSS-references: GB:M61737; GB:M62754; NID:g144835; PIDD:AAA3248.1; PID:g144836  
R:Peterson, J.W.; Lanzillo, W.N.; Lemon, B.J.; Seefeldt, L.C.  
Science 282, 1853-1858, 1998  
A>Title: X-ray crystal structure of the Fe-only hydrogenase (Cpl) from Clostridium pasteurianum.  
A:Reference number: A59203; MUID:99055388; PMID:9836629  
A:Contents: annotation; X-ray crystallography, 1.80 angstroms  
R:Peterson, J.W.; Lanzillo, W.N.; Lemon, B.J.; Seefeldt, L.C.  
submitted to the Protein Data Bank, October 1998  
A:Reference number: A77988; PDB:1FEH  
A:Contents: annotation; X-ray crystallography, 1.80 angstroms, residues 1-574  
C:Comment: Three distinct hydrogenases, the Fe, NiFe, and NiFeSe hydrogenases, have been identified, each consisting of a single polypeptide chain.

[illegible]

F:502/Binding site: diltion cofactor (Cys) #status predicted

Query Match. 28.8% Score 741; DB 2; Length 582;

Best Local Similarity 37.9% Pred. No. 1.8e-44;

Matches 167; Conservative 66; Mismatches 134; Indels 72; Gaps 9;

QY 52 VACAAAPAAEAPLSHVQALAEIAKPKDDPTKHKVGVAVAPAVAAIAETLGLAPATP 111  
 DB 197 IACPVAAALKKSHIEKQVQAL-----NDP-KKHVIAVAAAPSVRTAMGELFKMGYKDV 248  
 QY 112 PKLAEGRLRGDEVDFTLFCADLTIMEGSELHLRTELEAHPHSDEPLPMFTSCCP 171  
 DB 249 TGLLYALRMLGDKVDFINFGADMTIMEATELLGNV-----KNNPFPMTSCCP 300  
 QY 172 GWIAALEKSYPDILPYVSSCKSPQMLAAVVKYLAERKGIAPKDMVMSIMPCTROSE 231  
 DB 301 AWRLAQNHYPELIDNLSAKSPQOIFGTASKTYPSISGIAPEDEVYTYIMCNDKYE 360  
 QY 232 ADRDMFCVADPTLRQLDHYITVELGNIFKERGINLAELPEGEWNPMPGVSGAGVLF 291  
 DB 361 ADIPFMEYN---SLRIDASLTTRRELAKMIKDAKIFADEDEVDPAAGTYSGAGAF 417  
 QY 292 TTGGVMEALRTAYELFTGTPRLSLSEVRGMDGKETITVVPAPAGSFEEELKHRAA 351  
 DB 418 ATGVMEAAIRSAKDAENKLENDVTEVRGFKGKEAVEI-----460  
 QY 352 ARAEAAAHGTPPLAMDGAGFTSEDRGSGITLRVAVANGLGNAKKLITKMQAGEAKYDF 411  
 DB 461 -----AGNKLNAVINAGSNFFEFMKSGKNNKYOYH 492  
 QY 412 VEIMACPVAGVGGGQP--RSTDKA---ITQKQALYLNDEKSTL--RRSHENPSIREL 464  
 DB 493 IEVMACPGCGINGGQPHVNAIDRENDVYRKLASVLYNOD--KNVLSKRSHNPATIKM 551  
 QY 465 YDTYLGEPGLKHAHELTHY 485  
 DB 552 YDSIFGKPGEGLAHKLHVKY 572

# RESULT 7

D96903

hydrogenase dehydrogenase (imported) - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C:Accession: D96903

R:Moiling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: D96903

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-582 <KUR>

A:Cross-references: GB:AE00437; PIDN:AAK78015.1; PID:G15022048; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

A:Genetics:

A:Gene: CAC0028

C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology

Query Match 28.8% Score 741; DB 2; Length 582;  
 Best Local Similarity 37.9% Pred. No. 1.8e-44;  
 Matches 167; Conservative 66; Mismatches 134; Indels 72; Gaps 9;

DB 301 AWRLAQNHYPELIDNLSAKSPQOIFGTASKTYPSISGIAPEDEVYTYIMCNDKYE 360  
 QY 232 ADRDMFCVADPTLRQLDHYITVELGNIFKERGINLAELPEGEWNPMPGVSGAGVLF 291  
 DB 361 ADIPFMEYN---SLRIDASLTTRRELAKMIKDAKIFADEDEVDPAAGTYSGAGAF 417  
 QY 292 TTGGVMEALRTAYELFTGTPRLSLSEVRGMDGKETITVVPAPAGSFEEELKHRAA 351  
 DB 418 ATGVMEAAIRSAKDAENKLENDVTEVRGFKGKEAVEI-----460  
 QY 412 VEIMACPVAGVGGGQP--RSTDKA---ITQKQALYLNDEKSTL--RRSHENPSIREL 464  
 DB 493 IEVMACPGCGINGGQPHVNAIDRENDVYRKLASVLYNOD--KNVLSKRSHNPATIKM 551  
 QY 465 YDTYLGEPGLKHAHELTHY 485  
 DB 552 YDSIFGKPGEGLAHKLHVKY 572

# RESULT 8

HODVFL

cytochrome-c3 hydrogenase (EC 1.12.2.1) (Fe) large chain - Desulfovibrio vulgaris (st

N:Alternate names: hydrogenase (Fe) alpha chain

C:Species: Desulfovibrio vulgaris

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 05-May-2000

C:Accession: A24551; B27480

R:Voordouw, G.; Brenner, S.

Eur. J. Biochem. 148, 515-520, 1985

A:Title: Nucleotide sequence of the gene encoding the hydrogenase from Desulfovibrio

A:Reference number: A24551; MUID:85203856; PMID:3888621

A:Accession: A24551

A:Molecule type: DNA

A:Residues: 1-421 <VOO>

A:Cross-references: GB:X02416; NID:940829; PIDN:CA26266.1; PID:940830

R:Pickrill, B.C.; He, S.H.; Li, C.; Menon, N.; Choi, E.S.; Przybyla, A.E.; Derwentant

Biochem. Biophys. Res. Commun. 149, 369-377, 1987

A:Title: Identification of three classes of hydrogenase in the genus, Desulfovibrio.

A:Reference number: A27480; MUID:86106446; PMID:3322275

A:Accession: B27480

A:Molecule type: protein

A:Residues: 1-34, 'K' <PRI>

C:Comment: Hydrogenases catalyze reactions involving the production or consumption of

e, NlFe, and NlFeSe hydrogenases, are found in the periplasm of sulfate-reducing bact

C:Comment: The active D. vulgaris hydrogenase (Fe) is a dimer of large (alpha) and sm

ction of sulfate to hydrogen sulfide in an electron transport chain. Cytochrome c3 is

C:Genetics:

A:Gene: hysA

C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology

C:Key words: 4Fe-4S; hydrogen metabolism; iron-sulfur protein; metalloprotein; oxidore

F:2884/Domain: ferredoxin 2[4Fe-4S] homology <REP>

F:33,38,41,76/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:45,66,69,72/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:382/Binding site: diltion cofactor (Cys) #status predicted

Query Match 24.6% Score 633; DB 1; Length 421;  
 Best Local Similarity 36.6% Pred. No. 4.7e-37;  
 Matches 149; Conservative 56; Mismatches 122; Indels 80; Gaps 9;

QY 54 CAAAPAAEAPLSHVQALAEIAKPKDDPTKHKVGVAVAPAVAAIAETLGLAPATP 112  
 DB 72 CLTHCP--ENAIYAGSVVEVEKKLDGKVK--CIAMPAPAVYALGDAFGMVGSVTT 127  
 QY 113 KQLAEGRLRGDEVDFTLFCADLTIMEGSELHLRTELEAHPHSDEPLPMFTSCCP 172  
 DB 128 GKMLAALQIKGFACWCTEFTADVTIMEEGSEFEVERLTK-----KSDMPLPQFTSCCP 181  
 QY 173 WIAALEKSYPDILPYVSSCKSPQMLAAVVKYLAERKGIAPKDMVMSIMPCTROSEA 232

Db 182 WQVATYYPELLPHFSTCKSPIGNGALAKTYGAERMKYDPKOYTVASIMPCIAKKYEG 241  
QY 233 DEDWFCVADPTL-----RQDHYITTVTELGNIFERGINLAELPEGEMDNPMGVSGAG 287  
Db 242 LR-----PELKSSGRDIDATITTRRLAMIKKAGIDPFKLPDCKDLSMGESTGA 293  
QY 288 VLEFGTGGVMEALRTAYELFTGTPLPRLSLSEVRQMDIKETNITMVPAPGSKPEELK 347  
Db 294 TIFGTGGVMEALRTAYELFTGTPLPRLSLSEVRQMDIKETNITMVPAPGSKPEELK 340  
QY 348 HRAAARAEAAHGTGPEPLAMDGAGFTSEDRGCGITLRVAVANGCNAKKLITKMGAEA 407  
Db 341 -----GCTDVKAAVYHGAERFKQVCDVYAGRS 368  
QY 408 KYDFEIMACPGCVGGGQGP-----RSTDKAIRQ-----RQOAL 443  
Db 369 PHEFTEYMACPGCGVGGGQGPVPMGVLEMDRTTRLYAGLKKRLAM 415

RESULT 9  
HODVIV  
Cytochrome-c3 hydrogenase (EC 1.12.2.1) (Fe) large chain - Desulfovibrio vulgaris subsp.  
N. alternans names: hydrogenase (Fe) alpha chain  
C:Species: Desulfovibrio vulgaris subsp. oxamunic  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 05-May-2000  
C:Accession: A32886  
R:Voordouw, G.; Strang, J.D.; Wilson, F.R.  
J. Bacteriol. 171, 3881-3889, 1989  
A:Title: Organization of the genes encoding [Fe] hydrogenase in Desulfovibrio vulgaris  
A:Reference number: A32886; MUID:89291738; PMID:2661538  
A:Accession: A32886  
A:Molecule type: DNA  
A:Residues: 1-421 <VOO>  
A:Cross-references: GB:M27212; NID:9145098; PIDN:AAA33373.1; PID:9145099  
C:Comment: Hydrogenases catalyze reactions involving the production or consumption of H<sub>2</sub>, N<sub>2</sub>, and N<sub>2</sub>O. Hydrogenases, are found in the periplasm of sulfate-reducing bacteria  
C:Comment: The active D. vulgaris hydrogenase (Fe) is a dimer of large (alpha) and small  
C:Comment: sulfite to hydrogen sulfide in an electron transport chain. Cytochrome c3 is 11  
C:Genetics:  
A:Gene: hvdA  
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology  
C:Keywords: 4Fe-4S; hydrogen metabolism; iron-sulfur protein; metalloprotein; oxidoreduc  
F:28-84/Domain: ferredoxin 2[4Fe-4S] homology <FE>  
F:35,38,41/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:45,66,69/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:382/Binding site: diron cofactor (Cys) #status predicted

Query Match 23.3%; Score 599.5; DB 1; Length 421;  
Best local similarity 36.5%; Pred. No. 1.1e-34;  
Matches 135; Conservative 49; Mismatches 117; Indels 69; Gaps 5;

QY 87 WCVQVAPAVRAVIAETLGLAPGATTPKQALAEGLRLGDEVDLTGLADLTMEESSELL 146  
Db 102 VAMPAPAVRAVIAETLGLAPGATTPKQALAEGLRLGDEVDLTGLADLTMEESSELL 161  
QY 147 HRLTEHLAHPDSDEPLPMTSCCPQIMALEKSYDPLPYVSSCKSPQMLAAMVSKYL 206  
Db 162 QRLTKL-----DKLPQFTSCCPQIMALEKSYDPLPYVSSCKSPQMLAAMVSKYL 215  
QY 207 AEKGIAPKDMVAVSIMPTCKROSEADP--WFCVADAPTLRLDHLVITTVELGNIFER 264  
Db 216 ADRMKYDRAKYTVTSIMPTCKRYBGMRLQW-----DSGHMDIDITDITRLAYMIKKA 270  
QY 265 GINTLAELPGEWDPNGVSGAGVYFGTTGGVMEALRTAYELFTGTPLPRLSLSEVRQ 324  
Db 271 KIDFRLKDPGKRDTLMGESTGATLFGVGGVMEALRTAYELFTGTPLPRLSLSEVRQ 330  
QY 325 DKIKETNITMVPAPGSKPEELKRAAARAEAAHGTGPEPLAMDGAGFTSEDRGCGITL 384  
Db 331 QGVKEKTVV-----GVDV 345  
QY 385 RVAVANGCNAKKLITKMGAEAKYDFEIMACPGCVGGGQGP-----RSTDK 433

Db 346 KVAVHAGARRFDVCELVRKAPRAPHFIEFMACPGCGVGGGQGPVPMGVLEADRRSTBM 405  
QY 434 AITOKROAL 443  
Db 406 YAGLKKRLAM 415

RESULT 10  
G72256  
Hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Thermotoga maritima (stra  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: G72256  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: G72256  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-645 <ARN>  
A:Cross-references: GB:AE001794; GB:AE000512; NID:94981978; PIDN:AAJ36496.1; PID:9498  
A:Experimental source: strain MS8  
C:Genetics:  
A:Gene: TM1426  
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology  
C:Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase  
F:34,45,48,60/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
F:92,96,99,105/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type N1)  
F:143,146,149,196/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:153,186,189,192/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:295,350,482,486/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:486/Binding site: diron cofactor (Cys) #status predicted

Query Match 19.9%; Score 512.5; DB 2; Length 645;  
Best local similarity 30.2%; Pred. No. 2.6e-28;  
Matches 133; Conservative 56; Mismatches 159; Indels 93; Gaps 9;

QY 54 CAAAPADAPLISHVOALAEALPRDDPTKRKVCQVAPAVRAVIAETLGLAPGATTPK 113  
Db 192 CVAVCPGALSTINDIKLEALE-----SDKIVIGMIPAVRAVIAETLGLAPGATTPK 246  
QY 114 QLAEGRLGDFEVDLTGLADLTMEESSELLRLTEHLAHPDSDEPLPMTSCCPQIM 173  
Db 247 KIVSFLKTTGFDVDFVSGADLVAYEAEHFEYERLKK-----GERLPQFTSCCPAW 298  
QY 174 IAWLEKSYDPLPYVSSCKSPQMLAAMVSKYLAEKGLAPKDMVAVSIMPTCKROSEAD 233  
Db 299 VKHAERTYPOYLONLSSVSPQALGTVIKKIARLKYPEEKIFLYSPFCTAKKFEAL 358  
QY 234 RMFCVADPTLRLDHL-----VITTVELGNIFERGINLAELPEGEMDNPMGVSGAG 287  
Db 359 RE-----EHEGIVDITVTRRLAQILKMSRIDINVERPOQFDRPYGSSQAG 405  
QY 288 VLEFGTGGVMEALRTAYELFTGTPLPRLSLSEV---RGMDGKIKETNITMVPAPGSKFEE 344  
Db 406 LGFGKAGVGSCLVSLNE-----EIGIEKVDVKSPEGRIVAVETL----- 447  
QY 345 LKHAARAEAAHGTGPEPLAMDGAGFTSEDRGCGITLRVAVANGCNAKKLITKMGAEA 404  
Db 448 -----KQTSKGANIYGLGVKKFLERK- 472  
QY 405 GEAKYDFEIMACPGCVGGGQGPSTDKAITOKROALYINLDEKSTLRSHNSPIREL 464  
Db 473 ---DVEIIEVMACNCGVGGGQGPYPNDSRIREHRAKVLADTGISLTLTPVENLFLMKL 529  
QY 465 YDTYLGEPRLGKRAHEHLHTHY 485  
Db 530 YEEDLKE--HTRHEHLHTTY 548

```

RESULT 11
T40992
hyothetical protein SPCC1450.10c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40992
R:Lynne, M.; Rajadream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A:Reference number: 221962
A:Accession: T40992
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-538 <L>
A:Cross-references: EMBL:AL049559; PIDN:CA940177.1; GSPDB:GN00068; SPDB:SPCC1450.10c
A:Experimental source: strain 972H-; cosmid c1450
C:Genetics:
A:Gene: SPDB:SPCC1450.10c
A:Map position: 3
A:Introns: 31/3

Query Match
Best Local Similarity 24.2%; Score 311.5; DB 2; Length 538;
Matches 123; Conservative 84; Mismatches 179; Indels 123; Gaps 19;

OY 52 VACAAAAPAEADLSHVQALAEALAKRDPTRKHCVCQYAPAVAAIAETGLAGATT 111
DB 65 LACSGCTTSSEITLVNLSQYQEVILKLESRSQELIVYSLSPQVRAALAYGLSLQEIQ 124
OY 112 PKLAGSLRLGDFEVDPLFGADLTMEGSELLRLTEHLAAH----- 156
DB 125 AVLEWPTFKGLGFAHLDITNASREIVLQCAQEFCSWLOS-RAHKNOQVTSVYNEHP 183
OY 157 --PHS-----DEPLMFTSCCPGWIAMLEKSPDLIPYVSSCKSPQ 195
DB 184 LIPHSSTQISGVHSNTSSNGINENAVLPILSSCPMICVEKTHSNLIPNLSRVASQP 243
OY 196 MMLAAVKSTLAKEKGIAPDMVWSIMPTCKQSEADRMFCVADPTLRQLDHYITTV 255
DB 244 QACGRILKDWAVOQFSQRNDVWHLISLPCFKKLEASRDEF--SENGVADVSVLTPK 300
OY 256 ELGNIFPERGINLAELPEGEWMDPM-----GVSGAGVLFGTGG 295
DB 301 ELVEMKFLRIDPLELTK---NPDPQSTDAIPRYPRITTEBOIGS-----SSGG 349
OY 296 VMAALRTAYELFTGP--LPRLSLSEVRGMDGK-----ETN---ITVVPAPGSKFEE 344
DB 350 YMGVLSLYAAKMLFEGIDVGPYVMNKNK-DLREYTLRHPEINBOLISMATCYGFRNIQ 408
OY 345 LKHHRAAARAEAAHGTTPGLANDGAGFTSEDRGGITLRVAVANGLGNAKKLITKMQA 404
DB 409 NLVR-----VHG-----NSSVRKGRVLLKRVRSNAQNPTE----- 440
OY 405 GEAKYDFEIMACPGACVGGGGO-PRSTDKAITQKRQ-----AALYLLDKSTLRSHEN 458
DB 441 EPSNYDVEVMACGGGCGINGGQLPFVSERIVASARMMQOQVEKLY--EPGT--RSVDQ 496
OY 459 PSIRELDTYLGEPLGKRAHELLHTHYA 487
DB 497 SAVSYMLEQWVKDPT--LTPKFLHTSYRA 523

RESULT 12
S63206
LEU1 protein homolog YNL240c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hyothetical protein N1114
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
C:Accession: S63206; JQ0891; S67358; S72076
R:Pandolfo, D.; De Antoni, A.; Lanfranchi, G.; Valle, G.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63188
A:Accession: S63206
A:Molecule type: DNA

A:Residues: 1-491 <P>
A:Cross-references: EMBL:271516; NID:g1302273; PID:e239709; PID:g1302274; MIPS:YNL240
A:Experimental source: strain S288C
R:Nogee, I.; Johnston, M.
Gene 96, 161-169, 1990
A:Title: Isolation and characterization of the YNF1 gene of Saccharomyces cerevisiae,
A:Reference number: JQ0890; MID:91099671; PMID:2269430
A:Accession: JQ0891
A:Molecule type: DNA
A:Residues: 229-280, 'AIRA', 286-374, 'SSAE', 380-392, 'P', 396-463, 'PPA', 473-474, 'R', 476,
A:Cross-references: GB:M34709; NID:g171543; PIDN:AAA34618.1; PID:g171544
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.
submitted to the EMBL Data Library, February 1996
A:Description: DNA sequence of cosmid 14-5 from Chromosome XIV.
A:Reference number: S67355
A:Accession: S67358
A:Molecule type: DNA
A:Residues: 1-491 <P>
A:Cross-references: EMBL:269381; NID:g1183970; PID:e221952; PID:g1183974
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.
Yeast 12, 1071-1076, 1996
A:Title: The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open reading
A:Accession: S72076
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-491 <P>
A:Cross-references: EMBL:269381; NID:g1183970; PIDN:CAA93358.1; PID:g1183974
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996
C:Genetics:
A:Gene: SGD:NAK1
A:Cross-references: SGD:S0005184
A:Map position: 14L
A:Note: YNL240c

Query Match
Best Local Similarity 9.9%; Score 255; DB 2; Length 491;
Matches 114; Conservative 72; Mismatches 189; Indels 118; Gaps 18;

OY 52 VACAAAAPAEADLSHVQ-----QALAEALAKRDPTRKHCVCQYAPAVAAIAETGL 105
DB 60 LACSGCTTSSEITLVNLSQSHSVFLKNKSKLSQGD-----KFLVSVSPQRLSLAQYGL 115
OY 106 APGATTPKQALAEGLRGLFDEVDPLFGADLTME-----EGSEL-----LHRL 149
DB 116 -----TEAADLCIMNFEQKHFQCKYWGTEMGRITISKT 151
OY 150 TEHLEAHPHSDPL-----PMFTSCCPGWIAMLEKSPDLIPYVSSCKSPQMLAAVKS 204
DB 152 VERITIAKKOKENTGADRKLPLSAVCPGFLYTEKTRPOLVPMILNKSPOQITGSLIRA 211
OY 205 YLAKEKGIAPDMVWSIMPTCKQSEADRMFCVADPTLRQLDHYITVEIGNIFKER 264
DB 212 TLF-ESLAIAESFYHLSLMPCKLEASRPESIDOS-----IDCVITPREIVTMQEL 264
OY 265 GINLAEL-----PEGEWMDPMGVSGAGVLFGTGGVMAALRTAYELFTGT- 311
DB 265 NLDPKPSFLTBDTSLYGRLSPPG-WDPRVHASNIG--GTCGGV-----AVQYVAVQ 313
OY 312 ---PLPRLSLSEVRGMDGIGETNITVMPAPGSKFEEELTKR---AARAEAAHGTTPG 365
DB 314 RLHPGSGMIVLEGRNSDIV-----EYRLLDRIIAASLSGFRNIQNL 358
OY 366 AMDGAGFTSEDRGGITLRVAVANGLGNAKKLITKMQAGEARY--DFEIMACPGACV 423
DB 359 VKRLTSSGSEKRRNITALLKRRTPGRANRREMAATAATADYHSHYIEVNNCPGCMN 418
OY 424 GGGOPRSTDAITQKQALYLNDEKSTLRSHENPSIRLVPYVYGLGEPLGKRAHELLHT 483
DB 419 GGGLNLEQNSLKRKQLV-----QTLNKRH-GEELAVVDLTGLGPKLEAAARPLSL 469
OY 484 HYVAGVVEEKDK 496

```



Db 470 EYFAPVQAVEK 482

## RESULT 13

hydrogenase chain (ferredoxin) [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C:Accession: B97297

R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A:Reference number: A96900; MUID:2135925; PMID:1359325

A:Accession: B97297

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-450 &lt;KOR&gt;

A:Cross-references: GB:AE001437; PIDN:AAK81165.1; PID:q15026303; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3230

Query Match 9.5%; Score 245.5; DB 2; Length 450; Best Local Similarity 24.6%; Pred. No. 1.1e-09; Matches 96; Conservative 50; Mismatches 113; Indels 131; Gaps 14;

QY 87 VCQVAPAVRAIAETGLAEGATTPKQALBGRLEGEVDITFGAD-LTIME--ES 143

Db 155 VIAAFAPAIAGQFGENVSL-----EMLRFAFKKVGADVEAFAADMLTIKAEFEEN 207

QY 144 ELLHRLTELEAHPHSDPLPMFTSCCPGMIALEKSPDILPYVSSCKSPQMLAAVYK 203

Db 208 ELV-----NSKQDLMTSCCCPMMWVMIRKIKYDLARHVSPPSPITAGRYTK 256

QY 204 SYLAEEKIAPKDMVYVIMPTKROSEADRMFCVDADPTLRQDHYITVELGNIFKE 263

Db 257 KLPNCK-----VFEIPIAKKAES-RESQISDA-----IDFVLTFELNGIFDV 301

QY 264 RGINALELPEGEWMDPMGVSGAGVLEFGTGYM---EALR---TAVELTGTPPLPL 316

Db 302 LDIDELKPE--THTKYASREGRLYGTGVSTVDEAVRIRPNKHILFKSTK---- 354

QY 317 SLSEVGMGDIKETNITWVPAFGSKFEELKHRAARAFAAHGTPGLAMDGAGFTSE 376

Db 355 -----VDGYKD----- 360

QY 377 DGRGILTRVAVANGIAGNAKLLITKMGAGEAKYDFVEIMACPGCVGGGQPRSTDKAIT 436

Db 361 -----CKDILNKTQAGNIGANFLGKMGCVGGCVG-----PKAIV 395

QY 437 OKROA--ALXNDKSTLRSHENSIREL 464

Db 396 HKDGRSEVNTAESSEIKIYSDSRMDI 425

## RESULT 14

S31336

LEFT protein - yeast (Kluyveromyces marxianus var. lactis)

C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 29-Oct-1999

C:Accession: S31336

R:Mesolowski-Louvel, M.; Tanguy-Rougeau, C.; Fuhkhan, H.

submitted to the EMBL Data Library, February 1993

A:Reference number: S31336

A:Accession: S31336

A:Molecule type: DNA

A:Residues: 1-469 &lt;MES&gt;

A:Cross-references: EMBL:X70373; NID:95537; PIDN:CAA49833.1; PID:95538

A:Gene: LEFT

Query Match 9.0%; Score 232; DB 2; Length 469;

Best Local Similarity 23.9%; Pred. No. 1e-08; Matches 110; Conservative 62; Mismatches 167; Indels 122; Gaps 16;

QY 52 VACAAAPAAEAPLISHVOQALAEAK--PKDDPTKHCVOYAPAVRAIAETGLAPGA 109

Db 57 IACACITSSSEILLKSOHVKVLEKMSLELDERSLAVSISPQCRSLADYVSMC--- 113

QY 110 TTPKQALBELR-----RLGFDEVFDTLFRADITIMEGSELHRLTELEAHPHSD 161

Db 114 ----LADLRCPQNMKTKFNKAKYVVGQFGRSISIRNATLDRPEN----- 159

QY 162 PLPMFTSCCPGMIALEKSPDILPYVSSCKSPQMLAAVYSLAEKGIAPKDMVYV 221

Db 160 BEPLLCVSPCGVLVAEKRPPELLIHMIDYKSPQDITGLMKQ-----ADPCYHLS 211

QY 222 IMPCTRKQSEADRMFCVDADPTLRQDHYITVELGNIFKRGINL-----AEL 271

Db 212 IMPCTRKQSEADRMFCVDADPTLRQDHYITVELGNIFKRGINL-----AEL 271

QY 272 PEGWMDNPMGVSGAGVLEFGTGYVMEALR-----TAVELTFG 310

Db 265 CPSSGMDYKILHMLNBE-----SSSGYAYOTLSLQSSNPESDIITTEGKNSDVTETRLVK 321

QY 311 TPLPRLSLSEVRGMGDI-----KETNITWVPAFGSKFEELKHRAARAFAAHGTPGL 365

Db 322 SKGVYASSEVYGFNIQVLVRLKLSQSAVYKRGIK---VARRQSVLK----- 367

QY 366 AMDGAGFTSEDRGILTRVAVANGIAGNAKLLITKMGAGEAKYDFVEIMACPGCVGG 425

Db 368 -----SGTSE-----KTRVLT---ADPAKTFVEVMACPGSCING 402

QY 426 GQPRSTDKAITOKROALYNIDKSTLRSHENSIRELXD 466

Db 403 GL-LNEKMANRRKOLA-ODLSLAYTKVHSVNIPIVAYD 441

## RESULT 15

HODVFS

hydrogenase (EC 1.18.99.1) (fe) small chain precursor - Desulfovibrio vulgaris (strat

N/A:Alternate names: Fe hydrogenase beta chain

C:Species: Desulfovibrio vulgaris

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jun-2000

C:Accession: B24551; A27480

R:Voordouw, G.; Brenner, S.

Eur. J. Biochem. 148, 515-520, 1985

A:Title: Nucleotide sequence of the gene encoding the hydrogenase from Desulfovibrio

A:Reference number: A24551; MUID:85203856; PMID:3888621

A:Accession: B24551

A:Molecule type: DNA

A:Residues: 1-123 &lt;VOO&gt;

A:Cross-references: GB:X02416; NID:940829; PIDN:CAA26267.1; PID:940831

R:Prickall, B.C.; He, S.H.; Li, C.; Menon, N.; Choi, E.S.; Przybyla, A.E.; DerVartanl

Biochem. Biophys. Res. Commun. 149, 369-377, 1987

A:Title: Identification of three classes of hydrogenase in the genus, Desulfovibrio.

A:Reference number: A27480; MUID:88106446; PMID:3322275

A:Accession: A27480

A:Status: preliminary

A:Molecule type: protein

A:Residues: 35-69 &lt;PRI&gt;

C:Comment: Three distinctive types of hydrogenases, the Fe, NiFe, and NiFeSe, are fou

C:Comment: This hydrogenase complex contains three (Fe-4S) iron-sulfur clusters.

C:Genetics:

A:Gene: hvdB

C:Complex: heterodimer of large (alpha, see PIR:HODVFL) and small (beta) chains

A:Function: catalyze reactions involving the production or consumption of molecula

A:Note: may be involved in hydrogen uptake for the reduction of sulfate to hydroge

C:Superfamily: hydrogenase (fe) small chain

C:Keywords: hydrogen metabolism; iron-sulfur protein; oxidoreductase; periplasmic spa

Query Match 5.4%; Score 138.5; DB 1; Length 123;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 16:07:42 ; Search time 22 Seconds

(without alignments)  
936.988 Million cell updates/sec

Title: CAC80065

Perfect score: 2576

Sequence: 1 MSALYKPCAAVNSIRGSSCR.....HELHHTHYAGGVEEKDEKK 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	743.5	28.9	574	1 PHFL_CLOPA	P29166 Clostridium
2	633	24.6	421	1 PHFL_DESVH	P07598 desulfovibrio
3	599.5	23.3	421	1 PHFL_DESVH	P13629 desulfovibrio
4	255	9.9	491	1 NARL_YEAST	P23603 saccharomyces
5	232	9.0	469	1 LETI_KLUFA	P53998 kluyveromyces
6	138.5	5.4	123	1 PHFS_DESVH	P07603 desulfovibrio
7	130	5.0	124	1 PHFS_DESVH	P13628 desulfovibrio
8	112	4.3	776	1 HYPE_AZOVI	P40596 azotobacter
9	111	4.3	912	1 ICAS_RABIT	Q28730 cryotolagus
10	109.5	4.3	857	1 CLPB_ECOLI	P03815 escherichia
11	108.5	4.2	1887	1 PAS2_YEAST	P19097 s fatty acid
12	107.5	4.2	391	1 ALR_STRCO	O86786 streptomycete
13	106	4.1	1885	1 PAS2_CANAL	P43098 c fatty acid
14	105.5	4.1	3491	1 ERY1_SACER	Q03131 saccharopolis
15	105	4.1	447	1 ODP2_RHIME	Q91933 rhizobium
16	105	4.1	896	1 IF2_CHLMU	O9PKU0 chlamydia m
17	104	4.0	371	1 THIL_PARDE	P54610 paracoccus
18	103.5	4.0	377	1 PROB_METIRU	O9Z988 meliothermus
19	103.5	4.0	1799	1 LMB2_MOUSE	O61592 mus musculus
20	102	4.0	835	1 RIRI_PPKVA	P50633 pseudomonas
21	100	3.9	1842	1 PAS2_SCHPO	O10289 s fatty acid
22	99	3.8	479	1 DPO4_STRCO	O9A882 streptomycete
23	99	3.8	1322	1 PUR4_XYLPA	O9P4F6 xylophaga fas
24	98.5	3.8	395	1 PURK_STNP7	O54575 synochococcus
25	98.5	3.8	871	1 ARS6_NEUCR	P54898 n arg-6 pro
26	98.5	3.8	3695	1 LMA5_HUMAN	O15330 homo sapien
27	97	3.8	388	1 ALR2_AGRRT5	P58737 agrobacteri
28	97	3.8	454	1 UCR2_NEUCR	O60044 neurospora
29	96.5	3.7	732	1 YF48_HUMAN	O9H0M4 homo sapien
30	96	3.7	1181	1 PIR2_HUMAN	O00722 homo sapien
31	95.5	3.7	553	1 SPAN_RAT	O88506 rattus norv
32	95.5	3.7	686	1 EFG2_STRCO	O87844 streptomycete
33	95.5	3.7	810	1 SYFB_SYNY3	P74296 synochocyst

34	95	3.7	247	1 ATPD_SORBI	O07300 sorghum bic
35	95	3.7	511	1 DLDH_SCHPO	O00087 schizosacch
36	95	3.7	1857	1 PAS2_PENPA	P15368 p fatty acid
37	95	3.7	2541	1 TALI_HUMAN	O9Y490 homo sapien
38	95	3.7	2541	1 TALI_MOUSE	P26039 mus musculus
39	94.5	3.7	971	1 KBFL_MOUSE	P25799 mus musculus
40	94	3.6	748	1 CLAT_HUMAN	P28329 homo sapien
41	94	3.6	1487	1 ICP4_HSVK	P17473 equine hepr
42	94	3.6	2035	1 HPC1_HUMAN	P51610 homo sapien
43	93.5	3.6	394	1 ATOB_ECOLI	P76461 escherichia
44	93	3.6	563	1 ARAB_BACCHD	O9K093 bacillus ha
45	93	3.6	673	1 VID3_AGRRT5	P18593 agrobacteri

## ALIGNMENTS

RESULT 1  
ID PHFL\_CLOPA STANDARD: PRT: 574 AA.  
AC P29166;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Periplasmic [rel] hydrogenase 1 (EC 1.18.99.1).  
OS Clostridium pasteurianum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1501;  
RN [1]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
RC STRAIN=ATCC 6013 / W5.  
RX MEDLINE=92002005; PubMed=1911757;  
RA Meyer J., Gagnon J.;  
RT Primary structure of hydrogenase I from Clostridium pasteurianum.;  
RL Biochemistry 30:9697-9704(1991).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
RX MEDLINE=99055388; PubMed=9836629;  
RA Peters J.W., Lanzilotta W.N., Lemon B.J., Seefeldt L.C.;  
RT X-ray crystal structure of the Fe-only hydrogenase (Cpl) from  
RL Clostridium pasteurianum to 1.8-A resolution.;  
CC Science 282:1853-1858(1998).  
CC -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) = 2 oxidized  
CC ferredoxin + H(2).  
CC -1- CORRECTOR: THIS PROTEIN BINDS PROBABLY FOUR 4FE-4S CLUSTERS AND ONE  
CC H CLUSTER (POSSIBLE SIX-TO-EIGHT IRON CLUSTER).  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: Periplasmic.  
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF  
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.  
CC -1- SIMILARITY: TO THE D-VULGARIS (HIDENBROUGH) HYD-GAMMA GENE  
CC PRODUCT.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M81737; AAA23248.1; .  
CC PIR: A40330; HQCLIP.  
CC PDB: 1FEH; 06-JAN-99.  
CC InterPro: IPR001450; 4Fe4S\_ferredoxin.  
CC InterPro: IPR003149; Fe\_hyd\_SSU.  
CC InterPro: IPR004108; Fe\_hyd\_1g\_C.  
CC InterPro: IPR001041; Ferredoxin.  
CC Pfam: PF00037; fer4; 2.  
CC Pfam: PF00111; fer2; 1.  
CC Pfam: PF02256; Fe\_hyd\_SSU; 1.  
CC Pfam: PF02906; Fe\_hyd\_1g\_C; 1.

DR PROSITE: PS00198; 4FE4S\_FERREDOXIN; 2.  
 KW Oxidoreductase: Periplasmic; Iron-sulfur (4Fe-4S) (3D-structure.  
 FT METAL 34 34 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 46 46 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 49 49 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 62 62 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 98 98 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 101 101 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 107 107 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 147 147 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 150 150 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 153 153 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 157 157 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 190 190 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 193 193 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 196 196 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 200 200 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 299 299 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 300 300 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 355 355 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 499 499 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 503 503 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 SQ SEQUENCE 574 AA; 63828 MW; 17E28A74E23C7DEE CRC64;

Query Match 28.9%; Score 743.5; DB 1; Length 574;  
 Best Local Similarity 37.1%; Pred. No. 2.1e-44;  
 Matches 163; Conservative 65; Mismatches 142; Indels 69; Gaps 6;

QY 52 VCAAAAPAAEAPLSHVQOALAEIAKPKDDPRKHCYQVAPAVRAIAETLGLAPGATT 111  
 DB 198 IACPVAALSEKSHMDRVKAL-----NAPEKHVYAMAPSVASRASTELNMGEGVDV 249  
 QY 112 PQAALGLRLRFGDEVDTLFGADLTITMEGSELLRLTEHLEAPHSDEPLPMPTSCPP 171  
 DB 250 TGRITATLQGLQDFDKFDFINFGADWTIMEATELVQRI-----ENNRPMPMTSCPP 301  
 QY 172 GRIAMLEKSPDLIPYVSSCKSPOMLAAMVKSYLEAKGIAKPAKDMVMSIMPCRKOSE 231  
 DB 302 GAVROAENYYPETLNNLSSAKSPQIFGTASTKYTPSISGLDPKKNFTVYVMECTSKPE 361  
 QY 232 ADRDMCVADPTLRDLQVITTVELGNFKERGINLAELPGEENMPNGVSGAGVNLG 291  
 DB 362 ADRPOMEKDC---LRQIDAVITTRELAKKIKDKIPFALBLESEADPAAGEYSGAGALIG 418  
 QY 292 TTGVVMAALRTAYELFTGTPRLSLSEVRGMDGKIKETNITWVPAPGSKFEELKRAA 351  
 DB 419 ATGGVMAALRSKADKRAENAELEIDIEYKQVRGNGIKAEVET---NNKKY----- 466  
 QY 352 AAPEAAAHGTGGPLANDGAGFTSEDRGCTILRAVAVANGLNAAKLITRMOAGEAKYDF 411  
 DB 467 -----NNAVINGASMLFEFKMSGMINERKQYHF 493  
 QY 412 VEIMACPAGVGCGGPRSTDK-----AITOKROALYMLDEKSTLRSHENPSIELVD 466  
 DB 494 IEMVACHGCGVNGGGPHVNPDKLEVKYDIKKVRAVSLYMODELSLRKSHENTALVAKMQ 553  
 QY 467 TYLGEPLGKHAHELTHY 485  
 DB 554 NYFGKRGPRGAEHLHFXY 572

RESULT 2  
 PHFL\_DESVH STANDARD; PRT; 421 AA.  
 ID PHFL\_DESVH  
 AC P07598;  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 01-APR-1993 (Rel. 26, Last annotation update)  
 DE Periplasmic (Fe) hydrogenase large subunit (EC 1.18.99.1) (Fe  
 hydrogenase).  
 GN HYDA.  
 OS Desulfovibrio vulgaris (strain Hildenborough).  
 OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;

OC Desulfovibrio.  
 OX NCBI\_TaxID=882;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85203856; PubMed=3888621;  
 RA Voodou G., Brenner S.;  
 RT "Nucleotide sequence of the gene encoding the hydrogenase from  
 Desulfovibrio vulgaris (Hildenborough).";  
 RL Eur. J. Biochem. 148:515-520(1985).  
 RN [2]  
 RP SEQUENCE OF 1-35.  
 RX MEDLINE=88106446; PubMed=3322275;  
 RA Prickl B.C., He S.H., Li C., Menon N., Choi E.S., Przydyla A.E.,  
 RA Deravertan D.V., Peck H.D., Jr., Faugue G., Legall J., Teixeira M.,  
 RA Moura J., Moura J.G., Pail D., Huynh B.H.;  
 RT "Identification of three classes of hydrogenase in the genus,  
 Desulfovibrio.";  
 RL Biochem. Biophys. Res. Commun. 149:369-377(1987).  
 RN [3]  
 RP SEQUENCE OF 1-15 FROM N.A.  
 RA van den Berg W.A.M., Stokkermans J.P.W.G., van Dongen W.M.A.M.;  
 RL Submitted (Sep-1992) to the EMBL/GenBank/DBI databases.  
 CC -1- FUNCTION: MAY BE INVOLVED IN HYDROGEN UPTAKE FOR THE REDUCTION OF  
 CC SULFATE TO HYDROGEN SULFIDE IN AN ELECTRON TRANSPORT CHAIN.  
 CC CYTOCHROME C3 IS LIKELY TO BE THE PHYSIOLOGICAL ELECTRON CARRIER  
 CC FOR THE ENZYME.  
 CC -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) -> 2 oxidized  
 CC ferredoxin + H(2).  
 CC -1- COFACTOR: BINDS THREE 4FE-4S CLUSTERS. TWO OF THESE CLUSTERS ARE  
 CC PROBABLY COORDINATED TO THE FERREDOXIN LIKE DOMAIN IN THE N-  
 CC TERMINAL OF THIS PROTEIN.  
 CC -1- SUBUNIT: COMPOSED OF TWO SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- MISCELLANEOUS: [FE], [NIFE], AND [NIFES] HYDROGENASES APPEAR TO  
 CC REPRESENT THREE DISTINCT ENZYMES HAVING HYDROGENASE ACTIVITY.  
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF  
 CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X02416; CAA26266.1; -  
 CC EMBL: Z15142; CAA78848.1; -  
 CC PIR: A24551; HDVFL.  
 CC HSSP: P29166; IFEH.  
 DR InterPro: IP001450; 4Fe4S\_ferredoxin.  
 DR InterPro: IP004108; Fe\_hyd\_1g\_C.  
 DR Pfam: PF00037; fer4; 2.  
 DR Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
 DR PROSITE: PS00198; 4FE4S\_FERREDOXIN; 2.  
 KW Oxidoreductase; Electron transport; 4Fe-4S; Iron-sulfur;  
 Periplasmic.  
 FT METAL 35 35 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 38 38 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 41 41 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 45 45 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 66 66 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 69 69 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 72 72 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 76 76 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 77 77 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT CONFLICT 35 35 C -> K (IN REF. 2).  
 SQ SEQUENCE 421 AA; 45951 MW; 8E70A0775B6BAACD CRC64;

Query Match 24.6%; Score 633; DB 1; Length 421;  
 Best Local Similarity 36.6%; Pred. No. 6.9e-37;  
 Matches 149; Conservative 56; Mismatches 122; Indels 80; Gaps 9;

QY 54 CAAAPAAEAPLSHVQOALAEIAKPKDDPRKHCYQVAPAVRAIAETLGLAPGATT 112

```

DB 72 CITHCP--ENAIYEAQSWPEVEKIKDKVK--CIAMPAPAVRYALGDAFGMPYGSVTT 127
OY 113 KOLAGLRRLGDFEVDLTFGADLTIMESGSELLHLEHNPDSDEPLMENSQCCG 172
DB 128 KGMALALDGLGRHACHDDEFTADVTIMESSEVERLTK-----KSDMLPQFPISSCGF 181
OY 173 WTAMLEKSYDPLIPYVSSCKSPOMMLAAVKSYLEKKGIAPKDMVMSIMPCTRKOSBA 232
DB 182 WCKYMETYYPELLPHSTCKSPIGNMGALAKYTGARNMYKDYKROYTVYSIMCIACKYKG 241
OY 233 DRDMFCVADPFL-----RQLDHVTITVELGNIFKRGINLAEPLPEGEENDMPYVSGAG 287
DB 242 LR-----PELKSSGMDIDATLTTRRELAVMIKAGIDFALPDGKRDSLMSGESITGA 293
OY 288 VLEFGTGGVMEALRTAYELFTGTPLRLSLSEVRNMDCKENITNMPAPSGKFEELK 347
DB 294 TTFGTGGVMEALRTAYELFTGTPLRLSLSEVRNMDCKENITNMPAPSGKFEELK 340
OY 348 HRAAARAFAAAGTGGPLAMDGAGFTSEDRGGLRLRYAVANGLSNAKLITKMOGFA 407
DB 341 -----GGTIVKAYVHGAARFRQYCDVAKGS 368
OY 408 KIDPEYIMACPGACVGGGQ-----RSTDKATIQ-----KQAL 443
DB 369 PYHFTLYMACPGCGVCGGQPYMPGVLEAMDRTTTRLYAGLKKRLAM 415

RESULT 3
PHYL_DESVO
ID PHYL_DESVO STANDARD; PRT; 421 AA.
AC P13629;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Periplasmic [Fe] hydrogenase large subunit (EC 1.18.99.1) (Fe
GN HYDA.
OS Desulfovibrio vulgaris (strain Oxamticus Monticello).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=884;
RP SEQUENCE FROM N.A.
RX MEDLINE=89291738; PubMed=2661538;
RA Voordouw G., Strang J.D., Wilson F.R.;
RT "Organization of the genes encoding [Fe] hydrogenase in Desulfovibrio
RT vulgaris subsp. oxamticus Monticello."
RT J. Bacteriol. 171:3881-3889(1989).
CC -1- FUNCTION: MAY BE INVOLVED IN HYDROGEN UPTAKE FOR THE REDUCTION OF
CC SULFATE TO HYDROGEN SULFIDE IN AN ELECTRON TRANSPORT CHAIN.
CC CYTOCHROME C3 IS LIKELY TO BE THE PHYSIOLOGICAL ELECTRON CARRIER
CC FOR THE ENZYME.
CC -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) -> 2 oxidized
CC ferredoxin + H(2).
CC -1- COFACTOR: BINDS THREE 4FE-4S CLUSTERS. TWO OF THESE CLUSTERS ARE
CC PROBABLY COORDINATED TO THE FERREDOXIN LIKE DOMAIN IN THE N-
CC TERMINAL OF THIS PROTEIN.
CC -1- SUBUNIT: COMPOSED OF TWO SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- MISCELLANEOUS: [FE], [NIFE], AND [NIFESF] HYDROGENASES APPEAR TO
CC REPRESENT THREE DISTINCT ENZYMES HAVING HYDROGENASE ACTIVITY.
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: M27212; AAA3373.1;
DR PIR: A32866; HODVLY.
DR HSSP: P29166; 1FER.
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR004108; Fe_hyd_19_C.
DR Pfam: PF00037; fer4; 2.
DR Pfam: PF02906; Fe_hyd_19_C; 1.
DR PRINTS: PR00353; 4FE4SFEROXIN.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
KW Oxidoreductase; Electron transport; 4Fe-4S; Iron-sulfur;
KW Periplasmic.
KW
FT METAL 35
FT METAL 38
FT METAL 41
FT METAL 45
FT METAL 66
FT METAL 69
FT METAL 72
FT METAL 76
SQ SEQUENCE 421 AA; 46278 MW; 8E987ABC4DC7C965 CRC64;

Query Match 23.3%; Score 599.5; DB 1; Length 421;
Best Local Similarity 36.5%; Pred. No. 1,5e-34;
Matches 135; Conservative 49; Mismatches 117; Indels 69; Gaps 5;

DB 87 VCVQVAVAYVAIAETGLAPGATPQKOLAEGLRGLGFDEVPDPLFGADLTIMESSELL 146
DB 102 VIAMPAPAVRYALGDAFGPLVGYVTGKMFSALEKGFQDCHDNETADVTIMESSTERY 161
DB 147 HRLTEHLEAHPHSDEPLPMTSCCPGMIALEKSYDPLIPYVSSCKSPOMMLAAVKSYLE 206
DB 162 QRLTKKL-----DKRLPQPTSCPGMHVYVESLYELPRLHSSCKSPIGMLGTAKTYG 215
OY 207 AEKKGIAPKDMVMSIMPCTRKQSEADR--WFCVADPRLQLDHVTITVELGNIFKER 264
DB 216 ADMKRYDRAKVVYVSIMPCTAKKYEGMRDLM-----DSGNKDIDATITDRELAUYMKRA 270
OY 265 GINLAELPGEEMDNPNVGSAGVLFQTTGGVMEALRTAYELFTGTPLRLSLSEVRGM 324
DB 271 KIDFTKLPQGRKDTLNGESTGATLFGVGGVMEALRTAYAVNTKKRPSMPKVGRL 330
OY 325 DGIKETNITMVPAPGSKFELLKRAAARAFAAAGTGGPLAMDGAGFTSEDRGGLTL 384
DB 331 QGVKEATVNV-----GGVQV 345
OY 385 RVAVANGLSNAKLITKMOAGEAKYDPEIMACPGACVGGGQ-----RSTDK 433
DB 346 KVAVVGARFRFHDVCELYAKAPMHFIETMACPGCGCGQPYMPGVLEADRRSTRM 405
OY 434 AIQKQKQAL 443
DB 406 YAGLKKRLAM 415

RESULT 4
NARL_YEAST
ID NARL_YEAST STANDARD; PRT; 491 AA.
AC P23503;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear architecture related protein 1.
GN NARL OR YNL240C OR N114.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RX MEDLINE=97051596; PubMed=8896273;
RA Pandolfi D., de Antoni A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
RT reading frames including a novel gene encoding a globin-like

```

RT domain.\*;  
 RL Yeast 12:1071-1076(1996).  
 RN (2)  
 RP SEQUENCE OF 147-491 FROM N.A.  
 RX MEDLINE=91099671; PubMed=2269430;  
 RA Nogae I., Johnston M.;  
 RT "Isolation and characterization of the ZWF1 gene of Saccharomyces  
 cerevisiae, encoding glucose-6-phosphate dehydrogenase.\*"  
 RL Gene 96:161-169(1990).  
 CC -1- SIMILARITY: TO K.LACTIS LEFT AND S.POMBE SPPC1450.10C.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
 CC FRAMESHIFTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: 269381; CAA93358.1; -;  
 CC EMBL: 271516; CAA96145.1; -;  
 CC EMBL: M34709; AAA34618.1; ALT\_FRAME.  
 CC PIR: J00891; J00891.  
 CC SCD: S0005184; NARI.  
 CC InterPro: IPR004108; Fe\_hyd\_1g\_C.  
 CC Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
 SO SEQUENCE 491 AA; 54151 MW; FFE2A07A881B8C1 CRC64;

Query Match 9.9%; Score 255; DB 1; Length 491;  
 Best Local Similarity 23.1%; Pred. No. 1.6e-10;  
 Matches 114; Conservative 72; Mismatches 189; Indels 118; Gaps 18;

QY 52 VACAAAPAAEAPLASHVQ-----QALAEAKRKDDPTKRNHCVOVAPAVAAIAETLGL 105  
 DB 60 LACAGCITSSSEILTSKSHVFLKMKWGLSKQOOD---KELVAVSPQCRSLADQYGL 115  
 QY 106 APGATTPKQAEGLRLGFEDEVDTLFGADLTME-----EGSEL-----LHRL 149  
 DB 116 -----TLEADLCIMNFQKHFQCKYVGMGTLMGNLISKT 151  
 QY 150 TEHLEAHPHSDPEL-----PMFTSCCPGMIAMLEKSTPDLIPYSSCKSQMMLAMVKS 204  
 DB 152 VERITIAHKQKENTGADRKLPLSAVCPGFLYTEKRPOLVPMILNVKSPQOITGSLIRA 211  
 QY 205 YLAEEKGIAPKDMVMSIMPCTRKQSEADDMFCVADPPTLRQLDHYITVEIGNIFKER 264  
 DB 212 TPESLAIARESFYHLSIMPCFQKLEASRPESLDG-----IDCVITPREIVTMQLQL 264  
 QY 265 GINLAEL-----PEGEMDNPMGVSGAGVLEGTGGVMEALRTAYELFTGT- 311  
 DB 265 NLDFKSFLEDTSLYGLRSLPPG--WDPRVHMASNLG---GRCGY-----AYQVYAVQ 313  
 QY 312 ---PLPRLSLSEVRGMDGIEETNITWPAAGSKFEELIKR---MAARAAMAAHGTPGL 365  
 DB 314 RLHPGSMIVLEGRNSDIV-----EYRLLDHDDRIIAASELSGFRNIQNL 368  
 QY 366 AMDGAGFTSEDRGIGITLAVAVANGLNNAKLLITKMQAGEAKY--DFVEIMACPGAGCG 423  
 DB 359 VRKLTSGSGSERKRNITALKRKRTGPKANSREMAAATAADPYHSDYIEVNAACPGACMN 418  
 QY 424 GGGGPRSTDAITQKROALYMLDEKSTLRSHENPSIRELYDTYLGEPLGHAHELHAT 483  
 DB 419 GGGGLNSEQNSLRKQLY-----QTLNKRH--GEELAMVDDPLTLGPKLEAAARLSTL 469  
 QY 484 HYVAGVEEKDEK 496  
 DB 470 EYVFAPYKQAVK 482

RESULT 5  
 LEFT\_KLUOLA

ID LEFT\_KLUOLA STANDARD; PRT; 469 AA.  
 AC P53998;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE LEFT1 protein.  
 GN LEFT1.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 2359/152;  
 RA Meszowski-Louvet M., Tanguy-Rougeau C., Fukuhan H.;  
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO YEAST YNL240C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X70373; CAA49833.1; -;  
 CC InterPro: IPR004108; Fe\_hyd\_1g\_C.  
 CC Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
 SO SEQUENCE 469 AA; 52072 MW; 555DAAE7DEFB9FB7 CRC64;

Query Match 9.0%; Score 232; DB 1; Length 469;  
 Best Local Similarity 23.9%; Pred. No. 6.1e-09;  
 Matches 110; Conservative 62; Mismatches 167; Indels 122; Gaps 16;

QY 52 VACAAAPAAEAPLASHVQALAEAK--PRDDPTKRNHCVOVAPAVAAIAETLGLAPGA 109  
 DB 57 LACAGCITSSSEILTSKSHVFLKMKWGLSKQOOD---KELVAVSPQCRSLADQYGL 113  
 QY 110 TTPKQAEGLRLGFEDEVDTLFGADLTMEEGSELHRLTEHLEAHPHSDPE 161  
 DB 114 -----LADDRCRQNFQKTFNKNKYVGTGFGSISIRNATLKDQVPE 159  
 QY 162 PLPMFTSCCPGMIAMLEKSTPDLIPYSSCKSQMMLAMVKSYLEKKGIAPKDMVMS 221  
 DB 160 EGPFLCSVCPGFLYTEKRPOLVPMILNVKSPQOITGSLIRA-----ADPTCYHLS 211  
 QY 222 IMPCTRQSEADDMFCVADPPTLRQLDHYITVEIGNIFKERGINL-----AEL 271  
 DB 212 IMPCFDRKLEASR-----EECEKEDVCYITPKQVAMAGDLSIDKSTMYEYDSKEL 264  
 QY 272 PEGEMDNPMGVSGAGVLEGTGGVMEALR-----TAYELFTG 310  
 DB 265 CPFGMDYKLMHLSNEG---SSSGYAVQYLLSSQSNPESDITIEGKNSDVEYRLVSK 321  
 QY 311 TPLPRLSLSEVRGMDGIEETNITWPAAGSKFEELIKRRAAARAAMAAHGTPGL 365  
 DB 322 SKGVIASSSEVYFRNIQNLVRKLSQASVYKRGKIK---VKRGSQVLR----- 367  
 QY 366 AMDGAGFTSEDRGIGITLAVAVANGLNNAKLLITKMQAGEAYYDFEIMACPGAGCG 425  
 DB 368 -----SGETSE-----KTKYVLT--ADPATDVEVNAACPGSGCINGG 402  
 QY 426 GPRSTDAITQKROALYMLDEKSTLRSHENPSIRELD 466  
 DB 403 GL-LNEKNNANRRKQLA-QDLSLATYTVHVSVINPDYIHAYD 441

RESULT 6  
 PHFS\_DESVH  
 ID PHFS\_DESVH STANDARD; PRT; 123 AA.  
 AC P07603;  
 DT 01-APR-1988 (Rel. 07, Created)



RA Chen J.C., Mortenson L.E.;  
RT "Identification of six open reading frames from a region of the  
RT Azotobacter vinelandii genome likely involved in dihydrogen  
RT metabolism";  
RL Biochim. Biophys. Acta 1131:199-202(1992).  
CC  
CC -1- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HYPP FAMILY.  
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
CC ACYLPHOSPHATASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X63650; CAA45185.1; ALT\_INIT.  
DR EMBL; L23970; AAA19510.1; ALT\_INIT.  
DR InterPro: IPR001792; Acylphosphatase.  
DR InterPro: IPR004421; Hypp.  
DR InterPro: IPR000666; Sua5\_yjc10\_yrhc.  
DR Pfam: PF00708; Acylphosphatase; 1.  
DR Pfam: PF01300; Sua5\_yjc10\_yrhc; 1.  
DR ProDom: PD001884; Acylphosphatase; 1.  
DR ProDom: PD002209; Sua5\_yjc10\_yrhc; 1.  
DR Trifam: TRIGR00143; hypp; 1.  
DR PROSITE; PS00150; ACYLPHOSPHATASE\_1; 1.  
KW Zinc-finger.  
FT DOMAIN 1 93 ACYLPHOSPHATASE-LIKE.  
FT ZN\_FING 113 138 C4-TYPE (POTENTIAL).  
FT ZN\_FING 163 188 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 776 AA; 82732 MW; E979337C0A15600C CRC64;

Query Match 4.38; Score 112; DB 1; Length 776;  
Best Local Similarity 21.88; Pred. No. 2.5;  
Matches 130; Conservative 53; Mismatches 220; Indels 192; Gaps 28;

OY 29 PLASIVRVATLLEAPRLGN-VACAAAPAAEAPLS-----HV-----QQAAL 73  
DB 182 PIACPACGPRLCEDSGRSLSDPIERLALALRGELIALRGVGFHLCDARNAGAVA 241  
OY 74 ELAKPKDDPTKHCYQVAPAVVAIAE-TLGLA-PGATPKQIAEGIRLGFVEFD 129  
DB 242 ELRRRRRRPAKPPALMAANASIALVALVLENGLAELSGPAVVLLR-KRAADDLA 259  
OY 130 TLFGADITTEECSELIHRIERLEAHPHSDEPLPMFTSCPGWIAMLEKSYDLPYVS 189  
DB 300 AGVAPDLAWL-GAMLPHSPLHMLLFHEAAGR-----AGTQMA----- 337  
OY 190 SCSPQMLAAMKSYLAEEKGIAPKDMVAVSIMPCRKSEADR-----DMFCVDADP 243  
DB 338 ---APODLLVMTSANSIG---AP-----PTGNAREERLAGIDLMLHRET 361  
OY 244 TLRLDHTVITVELGNIFKERGINLAEPLGENDMPKVSAGVLEGTGGVMEALRT 303  
DB 382 LINCDDSVSALGRAPLVIRSGRLAPL-----EIPLRG-GRSIL--ALGG---QLKN 429  
OY 304 AYELEGTPLRLSLSEVRG-----MDGIKETNTMTVPAP----- 338  
DB 430 AVCLTRG---DRAMLSPHNGDLDDADTCRALERTVARLGLLIGRPERVACDLHPDFAS 486  
OY 339 -----GSKFEELKRAARAARAAHGTGCP---LAWG-GAGFTSEDS--RGIT 382  
DB 487 RPARDAVERHGLPLPHQHHAHIAATMAEHGLAEPLVGLALDGFGLG---ADGRUNGE 543  
OY 383 TLRYA-----VANGIGNAKKILTKMGAGEAKYDF 411  
DB 544 LRVVAADGCAWLGELNPLPLPGDDQASREPWRMAAGLHLNLGSGGELATRFAPAEPGAAT 603  
OY 412 VEIWA-----CP-----AGCVGGGGPSTDKAITO-----KQQAALYNL 446

DB 604 AMLERGFNCPPSSSAGRFLDAAAGLGLGGRRQRYEAFAALRLSESVGLPERVPALMRI 663  
OY 447 DEKSTLRSHENDPSINLEDTYTLGEPIGH-----KAHELTHRYVAGG 489  
DB 664 DDDNRDLPLPLARLADMDDTRAGAELEFGVLEALATMAARAATGTGHRVALG 718  
RESULT 9  
IC55\_RABIT  
ID ID IC55\_RABIT STANDARD; PRT; 912 AA.  
AC 028730;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Intercellular adhesion molecule-5 precursor (ICAM-5) (telencephalin).  
GN ICAM5 OR TICN.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_Taxid=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Japanese white; TISSUE=Brain;  
RX MEDLINE=94206534; PubMed=7794412;  
RA Yoshitake Y., Oka S., Nemoto Y., Watanabe Y., Nagata S.,  
RA Kagamiyama H., Mori K.;  
RT "An ICAM-related neuronal glycoprotein, telencephalin, with brain  
RT segment-specific expression.";  
RL Neuron 12:541-553(1994).  
CC  
CC -1- FUNCTION: ADHESION MOLECULE THAT BINDS TO LEUKOCYTE ADHESION  
CC LEA-1 PROTEIN (INTEGRIN ALPHA-L/BETA-2).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: Expressed on neurons in the most rostral  
CC segment of the mammalian brain, the telencephalon.  
CC -1- SIMILARITY: BELONGS TO THE IMUNOGLOBULIN SUPERFAMILY. ICAM  
CC SUPERFAMILY.  
CC  
CC -1- SIMILARITY: CONTAINS 9 IMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L31399; AAA18478.1; -  
DR HSSP; P05362; IIC1.  
DR InterPro: IPR003988; ICAM.  
DR InterPro: IPR003987; ICAM\_VCAM-1.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003598; Ig\_C2.  
DR InterPro: IPR003600; Ig\_Like.  
DR Pfam: PF00047; Ig; 6.  
DR PRINTS; PR01473; ICAM.  
DR PRINTS; PR01472; ICAMVCAM1.  
DR SMART; SM00410; IG\_Like; 3.  
DR SMART; SM00408; IGC2; 1.  
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Transmembrane;  
KW Repeat; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 912 INTERCELLULAR ADHESION MOLECULE-5.  
FT DOMAIN 30 826 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 827 847 POTENTIAL.  
FT DOMAIN 848 912 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 47 107 IG-LIKE C2-TYPE DOMAIN 1.  
FT DOMAIN 132 202 IG-LIKE C2-TYPE DOMAIN 2.  
FT DOMAIN 239 304 IG-LIKE C2-TYPE DOMAIN 3.  
FT DOMAIN 332 385 IG-LIKE C2-TYPE DOMAIN 4.  
FT DOMAIN 403 472 IG-LIKE C2-TYPE DOMAIN 5.  
FT DOMAIN 486 552 IG-LIKE C2-TYPE DOMAIN 6.  
FT DOMAIN 566 645 IG-LIKE C2-TYPE DOMAIN 7.



FT	DOMAIN	659	724	IG-LIKE C2-TYPE DOMAIN 8.
FT	DOMAIN	754	813	IG-LIKE C2-TYPE DOMAIN 9.
FT	DISULFID	54	101	BY SIMILARITY.
FT	DISULFID	139	195	BY SIMILARITY.
FT	DISULFID	246	297	BY SIMILARITY.
FT	DISULFID	339	378	BY SIMILARITY.
FT	DISULFID	430	465	BY SIMILARITY.
FT	DISULFID	493	546	BY SIMILARITY.
FT	DISULFID	573	638	BY SIMILARITY.
FT	DISULFID	666	717	BY SIMILARITY.
FT	DISULFID	761	806	BY SIMILARITY.
FT	CARBOHYD	53	53	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	192	192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	211	211	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	311	311	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	366	366	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	639	639	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	756	756	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	787	787	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	788	788	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	912 AA	95682 MM	4DC032853DA0EAE9 CRC64;
SO	SEQUENCE			
Query Match				
Best Local Similarity 4.3%; Score 111; DB 1; Length 912;				
Matches 114; Conservative 56; Mismatches 210; Indels 166; Gaps 26;				
OY	18	SCRANQVAPRAPLAATSTVALATLE----	APARRLGNVACAAAPAE-----	A 63
DB	138	SCRVPAGPGRSLTLLTLLGAGQELIRSFAGEPARAGAVLRTATVLAREDHGANSFCA		197
OY	64	PLSHVQQA-----	ELAKKDPTRKHNVCYAVARAIAETL-----	GL 105
DB	198	ELDLRPGGLAFENSSAPROLWTYALPLDSPR-----	LLAPVLEVDQSQSLVSCYLDGL	251
OY	106	ARGATTPKQALAGLRIGFDEVDFLFG-----	ADLTJMEEG-SELLHRLT-----	DHLEA 155
DB	252	FRASEGVIHALGDKRLN-PEV--TLEGAIYATATATAEBSIKVLVCATVIGGERRES		308
OY	156	HPR---SDPLPMFTSCPCGWTAMLEKSYDPLPYVSSCKSPQMWLAAMVKSYLEKKG1		212
DB	309	RENVTVYSPAPLLT-----	LSEPSAPEGKLVTVCTAGARALVTL-----	EGV 352
OY	213	AKKDAWVSIINPCTKQSPADRDPCVADPTLRQLDHYIT---	TYELGNIFERGINLA	269
DB	353	PAAPGQAPALQFNASESDGSRFC--	DATLELDGETLSKNGSAELRVLYAPR-LDDA	408
OY	270	ELPEGEWDPNG-----	VSGAGVLFCTGGVMEALRTAYELE	308
DB	409	DCPR-SWTPPEGPQTLRCBARGNPTPAVHACARSDGAVIALGLLPRTALAGTY----		463
OY	309	TGTPPLRLSLSEVRGMDGIKETNITVPAP-----	GSKEELLKRRAAAEAAAGTGP	363
DB	464	-----RCYAAANYG-EAVKQDVTLYEYAPALDSVCPERVTVLEGTBSLSCVANGVP		516
OY	364	PLAMDGAGFTSDGNGI-----	TLRAVAVANGLNNAKLITRMOAGEKTYFEV1	414
DB	517	PSV-SCVRFROADVLEGLLVAREHAGYRCBAINARALAKNAVATVEGPS----	FEE	570
OY	415	MACPAG-----	CVGGGG-----	OPRSTDKAITOK 438
DB	571	RSPFSWTVWVEGSEQLFSCGEVECKPPSPVQCVSGEASBGLLPLAPLNPSPDPSVPRD		630
OY	439	ROAALY 444		
DB	631	LAPGTY 636		

RESULT 10  
CLPB\_ECOLI  
ID CLPB\_ECOLI STANDARD; PRT; 857 AA.

AC P03815;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE ClpB protein (Heat shock protein F84.1).  
GN ClpB OR HTPW OR B2592 OR Z3886 OR ECS3455.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Psychrichia.  
NCBI\_Taxid=562, 83334;  
OX [1]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE-90239044; PubMed-2185473;  
RA Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M.,  
Matick J.S., Dallymple B., Kuramitsu H., Shiroza T., Foster T.,  
RA Clark W.P., Ross B., Squires C.L., Maurizi M.R.;  
RT "Conservation of the regulatory subunit for the Clp ATP-dependent  
RL protease in prokaryotes and eukaryotes.";  
Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517(1990).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
Science 277:1453-1474(1997).  
[3]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-97349890; PubMed-9205837;  
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakada S., Nakamura Y., Nashimoto H.,  
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,  
RA Yamaguchi H., Horinouchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RL analysis of its sequence features.";  
DNA Res. 4:91-113(1997).  
[4]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE-21074935; PubMed-11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimantata E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
Nature 409:529-533(2001).  
[5]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / RIMD 0509952;  
RX MEDLINE-21156231; PubMed-11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
DNA Res. 8:11-22(2001).  
[6]  
RN  
RP SEQUENCE OF 1-593 FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-91294165; PubMed-1906060;  
RA Kitagawa M., Wada C., Yoshiooka S., Yura T.;  
RT "Expression of ClpB, an analog of the ATP-dependent protease

RT regulatory subunit in *Escherichia coli*, is controlled by a heat shock  
 RT sigma factor (sigma 32).  
 RL J. Bacteriol. 173:4247-4253(1991).  
 RN [7]  
 RP SEQUENCE OF 753-857 FROM N.A.  
 RX MEDLINE-8247208; PubMed-6285294;  
 RA Shen W.-P., Squires C., Squires C.L.;  
 RT "Nucleotide sequence of the ring ribosomal RNA promoter region of  
 RL *Escherichia coli*."  
 RN Nucleic Acids Res. 10:3303-3313(1982).  
 RN [8]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RC STRAIN-K12 / W3110;  
 RA Ogura T., Tomoyasu T.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP SEQUENCE OF 1-14; 150-157; 355-364 AND 452-460.  
 RX MEDLINE-92062147; PubMed-1953774;  
 RA Pontis E., Sun X.Y., Joernvall H., Krook M., Reichard P.;  
 RT ClpB proteins copurify with the anaerobic *Escherichia coli*  
 RL reductase."  
 RN Biochem. Biophys. Res. Commun. 180:1222-1226(1991).  
 RN [10]  
 RP IDENTIFICATION AS A HEAT SHOCK PROTEIN.  
 RX MEDLINE-91294166; PubMed-2066329;  
 RA Squires C.L., Pedersen S., Ross B.M., Squires C.;  
 RT "ClpB is the *Escherichia coli* heat shock protein F84.1."  
 RL J. Bacteriol. 173:4254-4262(1991).  
 CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR  
 CC ATP-DEPENDENT PROTEASE.  
 CC -1- INDUCTION: BY HEAT SHOCK.  
 CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M29364; AAA24422.1; -;  
 DR EMBL: AE000345; AAC75641.1; -;  
 DR EMBL: D90887; BAA16476.1; -;  
 DR EMBL: AE005489; AAC57705.1; ALT\_INT.  
 DR EMBL: AP002562; BAB36878.1; -;  
 DR EMBL: X57620; CAA40846.1; -;  
 DR EMBL: V00350; CAA23639.1; -;  
 DR EMBL: U50134; AAA2959.1; -;  
 DR PIR: D35905; D35905.  
 DR PIR: S18736; S18736.  
 DR SWISS-2DPAGE: F03815; COLI.  
 DR ECO2DBASE: E072.0; 6TH EDITION.  
 DR ECO2DBASE: F084.1; 6TH EDITION.  
 DR Ecogene: EG10157; CLPB.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003959; AAA\_ATPase-centr.  
 DR InterPro: IPR001270; Chaprin\_n-clpA/B.  
 DR InterPro: IPR004176; CLP\_N.  
 DR Pfam: PF000004; AAA\_2.  
 DR Pfam: PF02861; CLP\_N; 2.  
 DR PRINTS: PR00300; CLP\_PROTASEA.  
 DR PRODOM: PD000739; GSP11\_E; 1.  
 DR SMART: SM00382; AAA; 2.  
 DR PROSITE: PS00870; CLPB\_2; 1.  
 DR PROSITE: PS00871; CLPB\_2; 1.  
 KW Chaperone: ATP-binding; Repeat; Heat shock; Alternative initiation;  
 KW Complete proteome.  
 FT CHAIN 1 857 CLPB PROTEIN.  
 FT INIT MET 149 857 CLPB PROTEIN 3.  
 FT DOMAIN 161 409 FOR CLPB PROTEIN 3.  
 FT DOMAIN 531 722 II.

FT NP\_BIND 206 213 ATP (POTENTIAL).  
 FT NP\_BIND 605 612 ATP (POTENTIAL).  
 FT CONFLICT 96 97 KL -> NV (IN REF. 1).  
 FT CONFLICT 122 122 L -> V (IN REF. 1).  
 SQ SEQUENCE 857 AA; 95585 MM; FD38CD96B2FC32A CRC64;  
 Query Match 4.38; Score 109.5; DB 1; Length 857;  
 Best Local Similarity 21.28; Pred. No. 4.2;  
 Matches 86; Conservative 67; Mismatches 162; Indels 91; Gaps 19;  
 QY 108 GATTPKQIACIAR-LGFDFEFDLFGADLTMESEGLHRLTLEHAPH---SDEPL 163  
 DB 312 GATTLDEYROYIERDAALERRFQVFAEPV-EDTIAILGKERYELHHRVQITPAI 370  
 QY 164 PMFTSCPGMIAMLEKSYV---DLIPYSSCKSPOM-----MLA 199  
 DB 371 VAAATLSHRITA--DROLPKAIDLIDEAASSTIMQDSKPEELDRIDRIQLKQQA 428  
 QY 200 AMVKSYLEAKKGIAPKDMVVSIMPTCKRQSEADRMFCVDAD---PTLR-QLDHVIT 254  
 DB 429 LMKSDASKKRL---DMLNELSDKEROYSELEEKAKKASLSGTOTIKAELEQAKIA 485  
 QY 255 VE---LGNIFKKGICLAEIPGEMDNPMGCVSGAGVLTGT---GGYMEALRTAY 305  
 DB 486 IEGARRVGDLARMSLEOYGRIPLEKOL---EAAQOLGKTRRLRNKVTDAEIAEVL 540  
 QY 306 ELFTGPTPLRLSSEVRGMGDIKETNTIWPAPGSKREELKRAARAEE----- 356  
 DB 541 ARWTGIVSRMSE-----REKLIR-----EOLHHRVICONENVDVSNAIR 585  
 QY 357 -AAHGPPLAMDGAGFTSEDRGCGITLRVAVANGLNKAKLTKMQAG--AKYDFVE 413  
 DB 586 RSRAGLADPNRPIGSPFLPPTGVKTELKALANFDESDMANVRIDMSEFMKHSYR 645  
 QY 414 IMACPACVCG---GGGPRSTDRAITKRAALYNLDEKSTLRSH 456  
 DB 646 LVGAPPGVYVEGGYLTLEA---VRRRYSVILLDE---VEKAH 683  
 RESULT 11  
 FAST\_YEAST STANDARD; PRT: 1887 AA.  
 ID P19097; Q12533;  
 AC 01-NOV-1990 (Rel. 16, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Fatty acid synthase subunit alpha (EC 2.3.1.86) [Includes: Acyl  
 DE carrier; 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)  
 DE (beta-ketoacyl reductase); 3-oxoacyl-[acyl-carrier protein] synthase  
 DE (EC 2.3.1.41) (Beta-ketoacyl synthase)].  
 GN FAS2 OR YPL231W OR P1409.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID:4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-88315020; PubMed-2900835;  
 RA Mohamed A.H., Chirala S.S., Mody N.H., Huang W.Y., Wakil S.J.;  
 RT "Primary structure of the multifunctional alpha subunit protein of  
 RT yeast fatty acid synthase derived from FAS2 gene sequence".  
 RL J. Biol. Chem. 263:12315-12325(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-X2180-1A;  
 RA Schueller H.J.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RX MEDLINE-97313271; PubMed-9169875;  
 RA Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansgorge W.,  
 RA Araujo R., Aparicio A., Bartelli B.G., Badcock K., Benes V.,

RA Bostein L., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delli H., Dipolo T., Dubois E., Ducrestoeft A.,  
RA Duncan M., Floeth U., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,  
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,  
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Matthe R., Messenguy F., Meves H.-W., Miripalli S., Moestl D.,  
RA Mueller-huer S., Nemath A., Neutlich U., Oefner P., Pearson D.,  
RA Patel F.X., Pohl T.M., Purnelle D., Schaefer M., Schaefe M.,  
RA Schreiers B., Schramm S., Schroeder M., Sidcu A.M., Tettelin H.,  
RA Urrestearzu L.A., Ushinsky S., Vierendeels F., Vissers S., Vosé H.,  
RA Walsh S.V., Mamunt R., Wang Y., Wedler E., Wedler H., Winnett E.,  
RA Zhong W.Y., Zollner A., Vo D.H., Hanf J.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";  
RL Nature 387:103-105(1997).  
RN [4]  
RP MOTAGENESIS OF GLY-1250.  
RC SPRAIN-5288C;  
RX MEDLINE-94316198; PubMed-8041367;  
RA Inokoshi J., Tomoda H., Hashimoto H., Watanabe A., Takeshima H.,  
RA Omura S.;  
RT "Cerulenin-resistant mutants of Saccharomyces cerevisiae with an  
RT altered fatty acid synthase gene.";  
RL Mol. Gen. Genet. 244:90-96(1994).  
CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF  
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.  
CC ALPHA SUBUNIT CONTAINS DOMAINS FOR: ACYL CARRIER PROTEIN,  
CC 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND 3-OXOACYL-[ACYL-  
CC CARRIER-PROTEIN] SYNTHASE. THIS SUBUNIT COORDINATES THE BINDING  
CC OF THE SIX BETA SUBUNITS TO THE ENZYME COMPLEX.  
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a  
CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).  
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-  
CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
CC [acyl-carrier protein].  
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +  
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
CC -1- SUBUNIT: [Alpha(6)beta(6)] hexamers of two multifunctional  
CC subunits (alpha and beta).  
CC -1- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM  
CC OTHER FUNGI.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-slb.ch/announce/>).  
CC Or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch).  
CC

DR EMBL; J03936; AAA34601.1; -;  
DR EMBL; X76890; CAAS4218.1; -;  
DR EMBL; X94561; CAA64256.1; -;  
DR EMBL; Z73586; CAA97947.1; -;  
DR EMBL; Z73587; CAA97948.1; -;  
DR PIR; A31107; A31107.  
DR SGD; S0006152; FAS2.  
DR InterPro; IPR002582; ACSB.  
DR InterPro; IPR000794; ketoacyl-synt.  
DR InterPro; IPR004568; Pantelhn.tn.  
DR InterPro; IPR003880; Pantelhn.tn.  
DR Pfam; PF00109; ketoacyl-syntc.1.  
DR Pfam; PF01648; ACSB.1.  
DR Pfam; PF02801; ketoacyl-synt\_C.1.  
DR ProDom; PD004282; ACSB.1.  
DR TIGRFAMs; TIGR00556; pantelhn.tn.1.  
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.  
DR PROSITE; PS00606; B-KETOACYL-SYNTHASE; 1.  
KW Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;  
KW Transferase; NADP; Phosphopantetheine.  
FT DOMAIN 1 ? ACYL CARRIER (ACP).  
FT 675 874 BETA-KETOACYL REDUCTASE.

FT	DNAI1	1149	1363	BETA-KETOACYL SYNTHASE.
FT	BINDING	180	180	PHOSPHOANETHEINE (BY SIMILARITY).
FT	ACT_SITE	1305	1305	BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT	MUTAGEN	1250	1250	G->S; CERULEIN-RESISTANCE.
FT	CONFLICT	310	310	G->GTTGG (IN REF. 1).
FT	CONFLICT	594	594	T->I (IN REF. 1).
FT	CONFLICT	941	1019	AKRLEIDTSEVRKAVSIETALAEKKVYNGSADAAYOVE TOPKANLVDEPELKPYQVOQAIAELEGGLDIERVI -> CLNCGYSWMLKLKLEOPRSKILMSIRLSMAIALMLHLKS KFNOELTFENMTSONNRHTRNLRLKLPLSLRWCMIKELF (IN REF. 1).
FT	CONFLICT	1036	1041	RMEWA->KMGNS (IN REF. 1).
FT	CONFLICT	1408	1408	A->S (IN REF. 1).
FT	CONFLICT	1671	1671	N->T (IN REF. 1).
SQ	SEQUENCE	1887 AA;	206946 MM;	08B872734CF3ABEA CRC64;
<hr/>				
Query Match		4.2%;	Score 108.5;	DB 1; Length 1887;
Best Local Similarity		23.2%;	Pred. No. 13;	
Matches		67;	Conservative 33;	Mismatches 116; Indels 73; Gaps 12
<hr/>				
OY	21 ARQVAPRAPIAASRYVALATLEAPARRIGNVACAAAPAAPAEAPLSHQALAEIAKPRD	80		
Db	104 AKKEAP-APTDAASAPAPAAAAPAV-----AAAPAA-----AAAEIA---D	142		
OY	81 DPTKHVCVOYAPAVRAVAIAETTLGLAPATTIPKOALEGRLPGEVEVDITFGADLTIME	140		
Db	143 EPVVASLLAHVA-----HKLRKSLDSTSPMSKTKRDVLVGKRSIV-183			
OY	141 EGSELHLHTLTHLEAHPHSDEPLP-----FTSCCPMIAMLEKSYPDLIPIYSSCKSP	194		
Db	184 --NELGDGEGFEFTTPKEPRETPLEIELAETFDQFFSG---ALGROSSLSRLISSRRP	238		
OY	195 QMMLAAMYKSYLAEEKGI--APKD---MVWVSIMPCTRROSEADRDFECVDADPTLROLD	249		
Db	229 GGFTTVARVKLTQRWGLPGSRQDGVLVALSNPEARAIGSPADAKAFI----DSMAQXY	294		
OY	250 HVTITVELGNIFERKGINLAELPEGEEMDNPKGVGSAGVFLPRTTGCVME	298		
Db	295 ASIVGVDDLSASAASAG-----MAGAGAAGAAIMIDAALAE	330		
<hr/>				
RESULT 12				
ALR_STRCO STANDARD: PRT; 391 AA.				
ID	ALR_STRCO	STANDARD:	PRT;	391 AA.
AC	086786;			
DT	30-MAY-2000 (Rel. 39; Created)			
DT	30-MAY-2000 (Rel. 39; Last sequence update)			
DT	15-JUN-2002 (Rel. 41; Last annotation update)			
DE	Alanine racemase (EC 5.1.1.1).			
GN	ALR OR SC04745 OR SC664_23.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Actinobacterias; Actinobacteria (class); Actinobacteridae;			
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=1902;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-AJ(2) / M145;			
RX	MEDLINE=21996410; PubMed=12000953;			
RA	Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Gobie A., Hidalgo J., Holmsby T., Howarth S., Huang C.-H., Kleiser T., Lathe L., Murphy L., Oliver K., O'Neill S., Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorek A., Woodward J., Barrett B.G., Parkhill J., Hopwood D.A.;			
RT	"Complete genome sequence of the model actinomycete Streptomyces coelicolor AJ(2).";			
RU	Nature 417:141-147(2002).			
CC	-1- FUNCTION: Provides the D-alanine required for cell wall biosynthesis (By similarity).			
CC	-1- CATALYTIC ACTIVITY: L-alanine = D-alanine.			

CC -1- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -1- PATHWAY: Along with D-alanine--D-alanine ligase, it makes up the  
 CC D-alanine branch of the peptidoglycan biosynthetic route.  
 CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AL031317; CAA20401.1; -  
 CC HSSP: P10724; 1BD0.  
 CC InterPro: IPR000821; Ala\_racemase.  
 CC Pfam: PF008642; Ala\_racemase; 1.  
 CC PRINTS: PR00992; ALARACEMASE.  
 CC TIGRPFAMS: TIGR00492; alt; 1.  
 CC PROSITE: PS00395; ALANINE\_RACEMASE; 1.  
 CC Isomerase: Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis;  
 KM Complete proteome.  
 FT ACT\_SITE 46 46 CATALYTIC BASE SPECIFIC TO D-ALANINE (BY  
 FT SIMILARITY).  
 FT ACT\_SITE 283 283 CATALYTIC BASE SPECIFIC TO L-ALANINE (BY  
 FT SIMILARITY).  
 FT BINDING 46 46 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT SEQUENCE 391 AA; 41199 MW; 24754AC1385DCA1B CRC64;  
 SQ  
 Query Match 4.28; Score 107.5; DB 1; Length 391;  
 Best Local Similarity 24.18; Pred. No. 2.2;  
 Matches 100; Conservative 50; Mismatches 130; Indels 135; Gaps 24;  
 Oy 5 VLKPCAIVSIRGSSCAROYAPRAPLAATVVALATLEAPARRIGVACAAAPAAEAP 64  
 Db 13 VLBARAEIDIALRANVRALRERAPGAA-----LMVVKADAGHGCHIPCARAVALAGAT 67  
 Oy 65 ---LSHVQALA-ELAKP-KDD-----PTRKHV--CYQVAPARVAIAETLG 104  
 Db 68 WLCTATPQGEALRLAAEPGLPDVRIKMWTPGFWREAVRLDVSAMAMEEVTG 127  
 Oy 105 LAGATTPKO-----LAEGIRLGFDEVPFTLGADLTTEEGSELLHRTLEHLEAH 156  
 Db 128 AAAAGVAPAVOLKADTGLRGCGCOPAD--WERLVGAAALRAEEGLRLTGLMSHAC- 184  
 Oy 157 PHSEDEPLPMTSCCPGMIAMLEKSYPLIRYSSCKSPQMLAAMVSYAEKKGIAPK- 215  
 Db 185 --ADEP-----GHPSI-----AAQITRRERKATV-AEORGLEFEV 216  
 Oy 216 -----DMV-----NVSIMPTCKRQSEAD--RDMECVADAPTL-RQL- 248  
 Db 217 RHIANSPATLTLDPADHDLVRPGIAMGVSPSEIGTPADFGRLRPVMTLAASLALVQVP 276  
 Oy 249 -----DHVITP-----VELG---NIFKERGINLALPGECMDNPG- 281  
 Db 277 GGHGVSTGHHYTPGETTIGLVPLGYADGIPRAHSSGPAVLVDGKMTVAAGRIAMQFV 336  
 Oy 282 -VGS-----GA-GVLFG-----TTGGVMEALRTAYELFT--GTPLPRLTISE 320  
 Db 337 DLGGDRPERGAENVLFGPDRGEPFTADMAQAAGTATVEITVNIIGSNVPRVYNE 391  
 RESULT 13  
 FAST2 CANAL STANDARD: PRT; 1885 AA.  
 AC P43098;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Fatty acid synthase subunit alpha (EC 2.3.1.86) [Includes: Acyl  
 DE carrier; 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)  
 DE (beta-ketoacyl reductase); 3-oxoacyl-[acyl-carrier protein] synthase  
 DE (EC 2.3.1.41) (beta-ketoacyl synthase)].

GN FAS2.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 ON NCBI\_TaxID=5476;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=4918;  
 RX MEDLINE=95255657; PubMed=7737507;  
 RA Southern S.B., Cihlar R.L.;  
 RT "Analysis and expression of the Candida albicans FAS2 gene.";  
 RL Gene 156:133-138(1995).  
 CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF  
 CC THE ALPHA SUBUNIT CONTAINS DOMAINS FOR: ALKYL-CARRIER PROTEIN,  
 CC 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND 3-OXOACYL-[ACYL-  
 CC CARRIER-PROTEIN] SYNTHASE.  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = 2  
 CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).  
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-  
 CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
 CC [acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +  
 CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
 CC -1- SUBUNIT: [alpha(6)beta(6)] hexamers of two multifunctional  
 CC subunits (alpha and beta).  
 CC -1- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM  
 CC OTHER FUNGI.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L29063; AAA34345.1; -  
 CC InterPro: IPR002582; ACPS.  
 CC InterPro: IPR000794; ketoacyl-synt.  
 CC InterPro: IPR004568; Pantethein\_tn.  
 CC InterPro: IPR003880; Pantethein\_attach.  
 CC Pfam: PF00109; ketoacyl-synt; 1.  
 CC Pfam: PF01648; ketoacyl-synt; 1.  
 CC DR Pfam: PF02801; ketoacyl-synt\_C; 1.  
 CC DR ProDom: PD004282; ACPS; 1.  
 CC DR TIGRPFAMS: TIGR00556; pantethn\_tn; 1.  
 CC DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.  
 CC DR PROSITE: PS00606; B\_KETOACYL-SYNTHASE; 1.  
 CC KM Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;  
 CC Transferase; NADP; Phosphopantetheine.  
 FT DOMAIN 1 2 ACYL CARRIER (ACP).  
 FT DOMAIN 2 2 BETA-KETOACYL REDUCTASE.  
 FT DOMAIN 3 3 BETA-KETOACYL SYNTHASE.  
 FT BINDING 181 181 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT ACT\_SITE 1304 1304 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
 SQ SEQUENCE 1885 AA; 207588 MW; 4835D57F362372E0 CRC64;  
 Query Match 4.18; Score 106; DB 1; Length 1885;  
 Best Local Similarity 21.66; Pred. No. 20;  
 Matches 63; Conservative 36; Mismatches 111; Indels 82; Gaps 13;  
 Oy 20 RANQVAPRAPLAATVVALATLEAPARRIGVACAAAPAAEAPLSHVQALAEALAKPK 79  
 Db 104 KOESTPSAPAAATPPPAANA---APV-----PAPAPASGVPSTI-----P 142  
 Oy 80 DDEPTRRHVCYQVAPAVVAIAETLGLAAGATTTPKQALGIRLGFDEVPFTLGADLTIM 139  
 Db 143 DEPVKANLLIHVVAQK--LKKPLDAVPMKAIKIDLVNGKSTV----- 183  
 Oy 140 EEGSELLHRTLEHLEAHPHSEDEPLP-----FTSCCPGMIAMLEKSYPLIRYSSCKS 193  
 Db 184 --ONEITIGDGKKEGFSPEKPEEDTPLEIAEPOFDSFSG--QLGKSTSLIGRLMSKM 238

DR	InterPro:IPR000794; Ketoacyl-synt.
DR	InterPro:IPR003880; Pantne_attach.
DR	Pfam: PF00106; adh_short; 1.
DR	Pfam: PF00109; ketoacyl-synt; 2.
DR	Pfam: PF00550; pp-binding; 3.
DR	Pfam: PF02801; ketoacyl-synt_C; 2.
DR	Pfam: PF02801; ketoacyl-synt_C; 2.
DR	PROSITE: PS00012; PHOSPHOPANTHETHEINE; 3.
DR	PROSITE: PS00056; B_KETOACYL_SYNTHASE; 2.
DR	PROSITE: PS00075; ACP_DOMAIN; 3.
KW	transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KM	Phosphopantetheine; Multifunctional enzyme.
FT	DOMAIN 1 1972
FT	DOMAIN 1979 3491
FT	DOMAIN 1 375
FT	DOMAIN 414 484
FT	DOMAIN 503 961
FT	DOMAIN 1030 1356
FT	DOMAIN 1611 1794
FT	DOMAIN 1888 1958
FT	DOMAIN 1979 2441
FT	DOMAIN 2507 2854
FT	DOMAIN 3055 3237
FT	DOMAIN 3334 3404
FT	ACT_SITE 145 145
FT	BINDING 447 447
FT	ACT_SITE 677 677
FT	ACT_SITE 1128 1128
FT	NP_BIND 1614 1660
FT	BINDING 1921 1921
FT	ACT_SITE 2148 2148
FT	ACT_SITE 2598 2598
FT	NP_BIND 3058 3104
FT	BINDING 3367 3367
SQ	SEQUENCE 3491 AA; 365022 MW; 682BFC32C90FA84 CRC64;
Query Match	4.18; Score 105.5; DB 1; Length 3491;
Best Local Similarity	21.68; Pred. No. 46;
Matches 117; Conservative	60; Mismatches 213; Indels 151; Gaps 25
OY	10 AAVSIRSSCARVAPRAPLAASTVRALATLEBA-----RLGNVCAAAAPAAPAEPL 65
DB	405 ASTGTRTRTCERLLA-----VNGETRALAGREDAENATRELGLDSVLAAQLAK--- 456
OY	66 SHVOAL--AELAKPKDDPTRKHVCOVADEVAYVALAETLG----- 104
DB	457 --VSALIGREVNIALLDHPTRLALAEALAGTAEQAQETBARTEARNEAPGEPAVVANAC 514
OY	105 -LAPGATTPKO----LAEGLRRL-----GFDEVFDTLFGADULTIMEGGSELHLRTHE 152
DB	515 RLPGGVSTPEFEFWELISGBRDVAVGLPTDRQMD--LDLSLFHPDP- -RSGT----- 561
OY	153 LEAHPSHDEPLPMFTSCOPGMIAMLKESYPDLPIVYSCKSPOMLAMVKSYLEAKKKGI 212
DB	562 --AHQRGGFTEATAFDPAFFGM-----SPREALVDPOORLMLELSNEYVERNGI 611
OY	213 APKDMVMWSIMPCTRKQSEADRMFCVDADPTLROLDIVITVELGINFIKEKGIMALELP 272
DB	612 PPTSL-----QASP-----TGVPFGLLPQETGRPLAE- 638
OY	273 EGEWDNPMVGSGAGVLFGTTGVMERALRTAYELFETGTPPLPRLSLEV----- 321
DB	639 -----GGEVEGYLMTGTTTSV-ASGRIVATYL-GLEGPAISVDPAACSSILVAHILA 687
OY	322 -----RGMDGKET-NITMWAPSPSKPEELLKHRAARAABEAHAICTPPPLAMDGAAGTTS 375
DB	688 CQSLRRRESSLTAMAGVTYMPPTPMLDVFSRNNSLAPDGRCRKAESA-----GANGGCM 740
OY	376 EDGGGGTLRLAVANGLNNAKLITTKMOAGEAKTVDEVI-NACPAGCVGGGGGPSPSTPKA 434
DB	741 ABEGGMLILE-RLSDARRNGHPVLAVLGRTAVNSDGSANGSLAP-----NGRAQVRVIDOA 795
OY	435 ITOK--ROAALYNIDKESTLRSHEBNPSIRELYDY--LGEP--LGKHAHELLHTHYA 487

